PI 1066594

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

September 17, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/402,275*
FILING DATE: *August 09, 2002*
RELATED PCT APPLICATION NUMBER: *PCT/US03/25084*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. Tarver

M. TARVER
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

D-1202275.080902 Approved

Attorney's Docket No. P0310
Express Mail Label No. EU582268497US

PATENT

COVER SHEET FOR FILING PROVISIONAL PATENT APPLICATION

Box Provisional Patent Application
Assistant Commissioner for Patents
Washington, DC 20231

This is a request for filing a PROVISIONAL PATENT APPLICATION under 37 C.F.R. 1.53(c).

Docket No.	P0310
Type a plus sign (+) inside this box →	+

INVENTOR(s)/APPLICANT(s)

First coinventor: Farid N. Najm, a U.S. Citizen residing in Mississauga, Ontario, Canada
Second coinventor: Navid Azizi, a Canadian Citizen residing in Markham, Ontario, Canada
Third coinventor: Andreas Moshovos, a Greek Citizen residing in Toronto, Ontario, Canada

TITLE OF THE INVENTION *(280 characters maximum)*

LOW LEAKAGE ASYMMETRIC SRAM CELL, ASSOCIATED NOVEL SENSE AMP, ASSOCIATED SRAM AND CACHE CELL STRUCTURES, AND RELATED METHODS

CORRESPONDENCE ADDRESS

Frank J. Pita
Registration No. 41,511
SRC
1101 Slater Road, Suite 120
Durham, NC 27703
Tel. Office (919) 941-9447
Fax Office (919) 941-9450

ENCLOSED APPLICATION PARTS *(check all that apply)*

- ☒ Specification (Number of Pages 8, including 8 Claims and Abstract) plus attachments:
- ☒ Attachment: The first paper attachment submitted herewith, entitled "*Low-Leakage Asymmetric-Cell SRAM*" prospectively being published August 12, 2002 at the ISLPED'02 conference in Monterey, California (4 Pages total, including Figures therein)
- ☒ Attachment: The second paper attachment submitted herewith, entitled "*Low-Leakage Asymmetric-Cell SRAM*" to be submitted to IEEE TVLSI (13 Pages total, including Figures)
- ☒ Drawing(s) (Included within Attachments listed above)
- ☒ PTO-2038 form for making Credit Card Payment of $160.00 large entity provisional patent application fee to U.S.P.T.O. (1 page total)
- ☒ Postage prepaid return receipt post card addressed to the above correspondence address
- ☒ Declarations and Powers of Attorney (for 2 of 3 inventors, 3 pages each, 9 pages total)

METHOD OF PAYMENT *(check one)*

☒ Payment by Credit Card using form PTO-2038 in the amount of $160.00 is enclosed to cover the provisional application filing fee ($160.00) for a large entity as required per 37 CFR 1.16(k).

☐ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account No. _____.

☐ Please charge Deposit Account No. _____ for any fee deficiency.

PROVISIONAL FILING FEE AMOUNT(s)

Large Entity $160.00
Small Entity $ 80.00

Filing Fee Amount: $160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted,

Frank J. Pita

Frank J. Pita
Registration No. 41,511

SRC
1101 Slater Road, Suite 120
Durham, NC 27703
Tel. Office (919) 941-9447
Fax Office (919) 941-9450

CERTIFICATE OF EXPRESS MAIL

"Express Mail" mailing label number EU582268497US
Date of Deposit: August 9, 2002

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Box Provisional Patent, Application, Assistant Commissioner for Patents, Washington, DC 20231.

Frank J. Pita

Frank J. Pita

LOW LEAKAGE ASYMMETRIC SRAM CELL, ASSOCIATED NOVEL SENSE AMP, ASSOCIATED SRAM AND CACHE CELL STRUCTURES, AND RELATED METHODS

FIELD OF THE INVENTION

The present invention relates generally to SRAM (Static Random Access Memory) structures and methods, and more particularly, to a low leakage power SRAM structures having low access latency, and related methods.

BACKGROUND OF THE INVENTION

See Section I. Introduction in each of two different versions of an attached paper titled "*Low-Leakage Asymmetric-Cell SRAM*" for the background of the invention. The first paper having the above title is prospectively being published August 12, 2002 at the ISLPED'02 conference in Monterey, California. The second paper having the above title is prospectively being submitted to an IEEE TVLSI editorial panel for prospective peer review and publication after the filing date of this provisional patent application. There is a need for improved SRAM structures having lower leakage current, and therefore lower leakage power requirements. There is a need for novel sense amp designs compatible with lower power SRAMs. There is also a need for lower power SRAM structures having access times comparable to those of non-power reduced SRAMs. There is a further need for better SRAM and SRAM cache structures, as well as for improved low power design methods.

SUMMARY OF THE INVENTION

The present invention provides a low leakage asymmetric SRAM cell, an associated novel sense amplifier, associated SRAM and cache cell structures, and methods related to the above that attempt to address at least some of the unmet needs discussed above.

The present invention provides an improved sense amplifier (sense amp) for the low leakage SRAM cell, as described in Section 3. Sense-Amplifier of the first paper noted above (ISLPED'02) and in Section III. Sense-Amplifier of the second paper noted above (IEEE TVLSI).

The present invention provides an improved low leakage SRAM, as described in Section 4. SRAM of the first paper noted above (ISLPED'02) and in Section IV. SRAM of the second paper noted above (IEEE TVLSI).

The present invention provides two improved low leakage cache organizations, as described in Section 5. Architectural Enhancements of the first paper noted above (ISLPED'02) and in Section V. Architectural Enhancements of the second paper noted above (IEEE TVLSI).

The present invention provides various low leakage power design methods as described in Section 2. through Section 6. of the first paper noted above (ISLPED'02) and in Section II through Section VI. of the second paper noted above (IEEE TVLSI).

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURES 1a through 10 are provided with appropriate captions in the first paper noted above (ISLPED'02).

FIGURES 1 through 23 and Tables I through XIV are provided with appropriate captions in the second paper noted above (IEEE TVLSI).

DETAILED DESCRIPTION OF THE INVENTION

The first paper attachment submitted herewith, entitled "*Low-Leakage Asymmetric-Cell SRAM*" prospectively being published August 12, 2002 at the ISLPED'02 conference in Monterey, California comprising four pages with figures as noted in the Brief Description of the Drawings is hereby incorporated by reference. The second paper attachment submitted herewith, entitled "*Low-Leakage Asymmetric-Cell SRAM*" prospectively being submitted to an IEEE TVLSI editorial panel for prospective peer review and publication after the filing date of this provisional patent application comprising thirteen pages with figures as noted in the Brief Description of the Drawings is hereby incorporated by reference.

Certain terms used within the first paper attachment and second paper attachments submitted herewith are defined herein. First, the IEEE is a professional association of electrical and electronic engineers known as the Institute of Electrical and Electronics Engineers. Further, the ISLPED is the International Symposium on Low-Power Electronics and Design. In addition, the TVLSI refers to Transactions on Very Large Scale Integration Systems.

A "high Vt transistor" as used in the paper attachment and in this patent application is defined as a transistor having a relatively higher "Vt" or threshold voltage than other transistors typically used in an SRAM cell design. The reason for selecting transistors having a higher Vt than others within the SRAM cell is to reduce the leakage current (and hence leakage power) in order to reduce an SRAM cell's leakage power. Although the high Vt transistor example described in the attached papers has a threshold voltage (Vt) which is 0.2volts higher, this is only an example for a 1.2volt, basic, .13 micron transistor. Different shifts of Vt could be used, using either higher or lower Vt differential voltages, so long as the leakage current draw is reduced as required for a given SRAM cell design or application.

An SRAM is a digital storage device composed of a number of SRAM cells, and as such each cell can store a binary variable, that can represent a binary "1" or "high" or a binary "0" or "low" value. A cache similarly is a digital storage device or memory device composed of a number of memory locations, such as a number of SRAM cells for

example, that may optionally be organized as an array. For example, a 64Kilobit cache may be organized as 8K storage locations each containing eight bits of information. For instance, a 64Kilobit cache may be organized as 64K storage locations each containing one bit of information.

A cache may be used in connection with a microprocessor. A cache may be implemented by using SRAMs or by using other memory types, such as DRAMs for instance. However, for speed reasons, SRAMs are the more common implementation vehicle for caches. Thus, SRAM cells may be used to build a cache. For example, a cache built out of SRAM cells may be referred to equivalently as "a cache" or as "an SRAM" or as "an SRAM array". These various terms are all commonly used in the art.

SRAM cells may also be used to build other types of memory structures that may not be employed or organized as a cache. For instance, one can build and use an SRAM chip, meaning a large array of cells comprising one whole chip that may be used in system design, to implement various types of memory. SRAM cells can also be used in FPGAs, ASICS, and other integrated circuits. SRAM cells can also be used other than as cache in microprocessor based designs, such as in register files or the like, for example.

A sense amplifier (sense amp) is a circuit comprised of several transistors, which is used to determine whether the content of a memory cell is a binary "1" or "high" or a binary "0" or "low" value, based on the directions in which the bitlines are being discharged. Any other terms not defined explicitly herein are to be construed according to the meaning ordinarily understood by those skilled in the art.

Where there is conflict between the description in the first and second papers and the description in this paragraph, the description in this paragraph should govern. There are two ways of storing the data in any SRAM array or cache- direct store and selective inversion. Direct store refers to storing data into a cache without inverting or changing all or part of a binary word (typically 8 bits or a byte, but not necessarily always this word size) before it is stored in the cache. Selective inversion refers to selectively inverting all or part of a binary word before it is stored in the cache. An "inversion flag" has to be stored to allow selective inversion. This inversion flag is to be implemented

(stored) using additional SRAM cells; these are details which should be familiar to those skilled in the art.

The present invention is described more fully herein with reference to the accompanying attached drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout this patent.

Although the foregoing invention has been described in some detail by way of the illustrations and examples provided in the paper attachments submitted herewith for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims. Those skilled in the art appreciate that alterations may be made to the above described invention without departing from the scope of the invention as claimed herein.

THAT WHICH IS CLAIMED IS:

1. An asymmetric SRAM cell for storing a binary variable having lower leakage power and access times comparable to conventional SRAM cells, comprising:

at least six transistors operably connected in an SRAM cell configuration, wherein a subset of the transistors are provided with a higher voltage threshold than the other transistors to lower leakage currents of the SRAM cell when the SRAM cell stores a binary variable representing a predetermined binary value.

2. A sense amplifier for an SRAM cell providing faster access times when the SRAM cell stores a first predetermined binary value, comprising:

four extra transistors, operably connected to ten transistors comprising a traditional sense amplifier for an SRAM cell;

four inputs, operably connected to a subset of the transistors in the sense amplifier; and

a set of dummy bitlines to trigger the reading of the predetermined binary value from the SRAM cell, wherein each pair of dummy bitlines are tied to dummy cells which store a second predetermined binary value, such that during every read access of the SRAM cell, one of the dummy cells will have its wordline asserted.

3. An array of asymmetric SRAM cells, wherein each SRAM cell stores a binary variable, the array organized as an SRAM device.

4. An array of asymmetric SRAM cells, wherein each SRAM cell stores a binary variable, the array organized as a cache device selected from the group consisting of a direct store cache and a selectively inverted cache.

5. An array of asymmetric SRAM cells, wherein each SRAM cell stores a binary variable, the array organized as an SRAM array selected from the group consisting of a direct store SRAM array and a selectively inverted SRAM array.

6. A method of designing a lower leakage power SRAM cell comprising the step of selecting transistors within the SRAM cell that are provided with a higher Vt than other transistors.

7. A method of operating a sense amplifier comprising the steps of precharging all four sense amplifier inputs, discharging an SRAM cell's bitline, and reading the SRAM cell after suitably discharging the bitline.

8. A method of accessing an SRAM cell using a sense amplifier comprising the steps of providing a sense amplifier having at least four inputs, accessing dummy cells, and accessing dummy bitlines.

ABSTRACT OF THE DISCLOSURE

The present invention provides a low leakage asymmetric SRAM cell, an associated novel sense amplifier, associated SRAM and cache cell structures, and methods related to the above. The present invention provides an improved sense amplifier for the low leakage SRAM cell. The present invention provides an improved low leakage SRAM. The present invention provides two improved low leakage cache organizations. The present invention provides various low leakage power design methods, sense amplifier and SRAM cell operation methods, and other methods.

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

Attorney Docket No. P0310

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am an original, first and joint inventor (of the three named coinventors listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

LOW LEAKAGE ASYMMETRIC SRAM CELL, ASSOCIATED NOVEL SENSE AMP, ASSOCIATED SRAM AND CACHE CELL STRUCTURES, AND RELATED METHODS

the specification of which
☒ is attached hereto
OR
☐ was filed on as United States Application No. or PCT International Application Number
and was amended on (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37 Code of Federal Regulations, § 1.56.

All known joint inventors for the above titled invention filing separate declarations and powers of attorney are:

First coinventor: Farid N. Najm, a U.S. Citizen residing in Mississauga, Ontario, Canada
Second coinventor: Navid Azizi, a Canadian Citizen residing in Markham, Ontario, Canada
Third coinventor: Andreas Moshovos, a Greek Citizen residing in Toronto, Ontario, Canada

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below any foreign application for patent or inventor's certificate, or of any PCT International application having a filing date before that of the application on which priority is claimed.

NONE			
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below.

NONE	
Application Number	Filing Date (MM/DD/YYYY)
Application Number	Filing Date (MM/DD/YYYY)

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) or § 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application (37 C.F.R. § 1.63(d)).

NONE		
Application No.	Filing Date	Status Patented/Pending/Abandoned
Application. No.	Filing Date	Status Patented/Pending/Abandoned
Application No.	Filing Date	Status Patented/Pending/Abandoned

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the practitioners associated with the Registration Number provided below to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith, and direct that all correspondence be addressed to that Registration Number:

Direct correspondence to the attendon of and telephone calls to:

Frank J. Pita
Registration No. 41,511
SRC
1101 Slater Road, Suite 120
Durham, NC 27703
Tel Office (919) 941-9447
Fax Office (919) 941-9450

Full name of second coinventor: Navid Azizi
Inventor's Signature: N.A. Date: 2002/08/08
Residence: ~~Toronto, Ontario, Canada~~ Markham, Ontario, Canada
Citizenship: Canadian
Post Office Address: 5 Montgomery Court, Markham, Ontario, Canada L3R 0C4

P0310

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

Attorney Docket No. P0310

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am an original, first and joint inventor (of the three named coinventors listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

LOW LEAKAGE ASYMMETRIC SRAM CELL, ASSOCIATED NOVEL SENSE AMP, ASSOCIATED SRAM AND CACHE CELL STRUCTURES, AND RELATED METHODS

the specification of which
☒ is attached hereto
OR
☐ was filed on as United States Application No. or PCT International Application Number
and was amended on (if applicable).

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37 Code of Federal Regulations, § 1.56.

All known joint inventors for the above titled invention filing separate declarations and powers of attorney are:

First coinventor: Farid N. Najm, a U.S. Citizen residing in Mississauga, Ontario, Canada
Second coinventor: Navid Azizi, a Canadian Citizen residing in Markham, Ontario, Canada
Third coinventor: Andreas Moshovos, a Greek Citizen residing in Toronto, Ontario, Canada

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below any foreign application for patent or inventor's certificate, or of any PCT International application having a filing date before that of the application on which priority is claimed.

NONE			
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed
Number	Country	MM/DD/YYYY Filed	☐ Yes ☐ No Priority Claimed

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below.

NONE	
Application Number	Filing Date (MM/DD/YYYY)
Application Number	Filing Date (MM/DD/YYYY)

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s) or § 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application (37 C.F.R. § 1.63(d)).

NONE		
Application No.	Filing Date	Status Patented/Pending/Abandoned
Application. No.	Filing Date	Status Patented/Pending/Abandoned
Application No.	Filing Date	Status Patented/Pending/Abandoned

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the practitioners associated with the Registration Number provided below to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith, and direct that all correspondence be addressed to that Registration Number:

Direct correspondence to the attention of and telephone calls to:

Frank J. Pita
Registration No. 41,511
SRC
1101 Slater Road, Suite 120
Durham, NC 27703
Tel Office (919) 941-9447
Fax Office (919) 941-9450

Full name of first coinventor: Farid N. Najm
Inventor's Signature: _____ Date: Aug. 8, 2002
Residence: Mississauga, Ontario, Canada
Citizenship: United States of America
Post Office Address: 3126 Workman Drive
Mississauga, Ontario
Canada L5M 6K5

P0310

Low-Leakage Asymmetric-Cell SRAM[1]

Navid Azizi
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
nazizi@eecg.toronto.edu

Andreas Moshovos
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
moshovos@eecg.toronto.edu

Farid N. Najm
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
f.najm@utoronto.ca

ABSTRACT

We introduce a novel family of asymmetric dual-V_t SRAM cell designs that reduce leakage power in caches while maintaining low access latency. Our designs exploit the strong bias towards zero at the bit level exhibited by the memory value stream of ordinary programs. Compared to conventional symmetric high-performance cells, our cells offer significant leakage reduction in the zero state and in some cases also in the one state albeit to a lesser extend. A novel sense-amplifier, in coordination with dummy bitlines, allows for read times to be on par with conventional symmetric cells. With one cell design, leakage is reduced by 7X (in the zero state) with no performance degradation. An alternative cell design reduces leakage by 40X (in the zero state) with a performance degradation of 5%.

Categories and Subject Descriptors

B.3.1 [Memory Structures]: Semiconductor memories

General Terms

Design

Keywords

SRAM, Low-leakage, Low-power, Dual-V_t

1. INTRODUCTION

As a result of technology trends, leakage (static) power dissipation has emerged as a first-class design consideration in high-performance processor design. Historically, architectural innovations for improving performance relied on exploiting ever larger numbers of transistors operating at higher frequencies. To keep the resulting switching power dissipation at bay, successive technology generations have relied on reducing the supply voltage. In order to maintain performance, however, this has required a corresponding reduction in the transistor threshold voltage. Since the MOSFET sub-threshold leakage current increases exponentially with a reduced threshold voltage, leakage power dissipation has grown to be a significant fraction of overall chip power dissipation in modern, deep-submicron ($< 0.18\mu$) processes. Moreover, it is expected to grow by a factor of five every chip generation [1]. For processors, it is estimated that in 0.10μ technology, leakage power will account for about 50% of the total chip power [2].

Since leakage power is proportional to the number of on-chip transistors, much of recent work in reducing leakage power has focused on SRAM structures such as the caches that comprise the vast majority of on-chip transistors. Existing circuit-level leakage reduction techniques are oblivious to program behavior and trade off performance for reduced leakage where possible [3]. Combined circuit- and architecture-level techniques reduce leakage for those parts of the on-chip caches that remain unused for long periods of time (thousands of cycles). These methods are not effective when most of the cache is actively used.

We present a family of novel *asymmetric* SRAM cell designs that lead to new cache designs which we refer to as the *Asymmetric-Cell Caches* (ACCs). ACCs offer drastically reduced leakage power compared to conventional caches even when there are few parts of the cache that are left unused. ACCs exploit the fact that in ordinary programs most of the *bits* in caches are *zeroes* for both the data and instruction streams. It has been shown that this behavior persists for a variety of programs under different assumptions about cache sizes, organization and instruction set architectures, even when assuming perfect knowledge of which cache parts will be left unused for long periods of time [4].

Traditional SRAM cells are symmetrically composed of transistors with identical leakage and threshold characteristics. Some recently proposed SRAM cells use symmetric configurations of transistors with different leakage and threshold characteristics [3]. These cells are either optimized for access latency or leakage power but not both. Our *asymmetric* SRAM cell designs offer low leakage with little or no impact on latency. In our asymmetric SRAM cells, selected sets of transistors are "weakened" to reduce leakage when the cell is storing a zero (the common case). In this work, we achieve the weakening by using higher-V_t transistors, however, this may also be possible by appropriate transistor sizing. We evaluate our designs by simulation, based on a commercial 0.13μ, 1.2V CMOS technology. The two best designs offer different performance/leakage characteristics. With one cell design, leakage is reduced by 7X (in the zero state) with no performance degradation. An alternative cell design reduces leakage by 40X (in the zero state)

[1]This project was supported in part by the Semiconductor Research Corporation (SRC 2001-HJ-901).

ISLPED'02, August 12-14, 2002, Monterey, California, USA

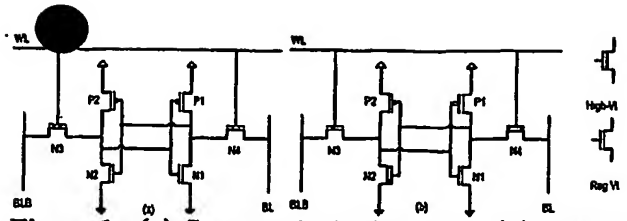

Figure 1: (a) Symmetric SRAM cell. (b) Original Asymmetric (OA) SRAM cell

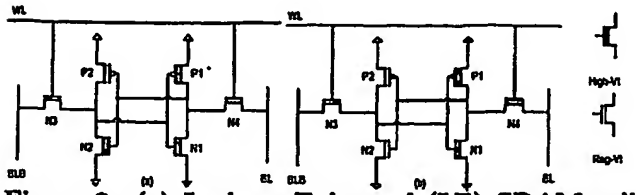

Figure 2: (a) Leakage Enhanced (LE) SRAM cell. (b) Speed Enhanced (SE) SRAM cell

with a sense time degradation of 10% (the total read cycle time is degraded by only 5%). By comparison, the use of an all high-V_t (HV) cell reduces leakage by about 40X but increases sense times by 26%.

We make the following contributions: (1) We introduce a novel family of asymmetric SRAM cells. No previous work on designing asymmetric SRAM cells exists. (2) We introduce a novel sense amp design that exploits the asymmetric nature of our cells to offer cell read times that are on par with conventional symmetric SRAM cells. (3) We evaluate a cache design that is based on ACCs and demonstrate that compared to a conventional cache, it offers drastic leakage reduction while maintaining high performance and comparable noise margins and stability.

The rest of this paper is organized as follows: In section 2, we present our asymmetric cell family. In section 3, we present the sense amplifier. In section 4, we present the simulation results of an SRAM using the different asymmetric cells. Section 5 includes a discussion on architectural level techniques to leverage the asymmetric nature of the cells. Finally, we conclude the paper in section 6.

2. ASYMMETRIC SRAM CELLS

Fig. 1(a) shows a conventional SRAM cell. In the inactive state, when the cell is not being written to or read from, most of the leakage power is dissipated by the transistors that are *i)* off and that *ii)* have a voltage differential across their drain and source. In Fig. 1(a), if the cell were storing a '0', transistors P1, N2 and N4 would dissipate leakage power. A simple technique for reducing leakage power would be to replace all transistors with high-V_t ones, but this unacceptably degrades the bitlines discharge times.

Since ordinary programs exhibit a strong bias in cache-resident bit values [5], another possibility to reduce leakage power, but at the same time keep read access times short, is to choose a preferred stored value and to only replace those transistors that contribute to the leakage power in this state with high-V_t transistors, as seen in Fig. 1(b).

This original asymmetric cell (OA) cell was simulated (at 110°C using SPICE models of a commercial 0.13μ, 1.2V CMOS technology and it exhibited the same leakage as the all regular-V_t (RV) cell when holding a logical '1', but it decreased leakage by 40X when holding a logical '0.' *Throughout this paper, we will use the following convention.* A high-V_t (HV) transistor is obtained from the basic 0.13μ, 1.2V, transistor (referred to herein as the *regular-V_t* transistor) by artificially increasing the V_t by 0.2V using the HSPICE in-line parameter DELVTO. It is understood that one may question the specific choice of 0.2V in practice. However, one can argue that the conclusions of this work, namely the feasibility and utility of using an asymmetric cell to reduce leakage, are valid irrespective of the specific value used for DELVTO. This specific value was selected, in our case, because it leads to a difference of about 10X between the leakage currents of HV and RV transistors, which is typical of dual-V_t technology.

The read access time of this OA cell is degraded. Due to N2's and N4's higher threshold voltage, they increase the bitline discharge time. The discharge times for the BLB and BL are 12.2% and 46.4% longer than the discharge times for the RV cell respectively. Read times can be made to match the faster read time by using a set of dummy bitlines and a novel sense amp, as is discussed in Section 3.

2.1 Two Improved Asymmetric SRAM Cells

Starting with the asymmetric cell of Fig. 1(b), we have investigated a total of 9 meaningful variations that offer different leakage and performance characteristics. In the interest of space, we present the two best designs, which are shown in Fig. 2.

The leakage enhanced (LE) cell in Fig. 2(a) offers better leakage behavior than that of Fig. 1(b) because it also dissipates reduced power when holding a logical '1' since N1 and P2 have been made high V_t. Compared to the RV cell it decreases leakage by 40X and 7X when holding a logical '0' and '1' respectively. The discharge times for this cell are 12.2% and 61.2% longer on BLB and BL respectively compared to the RV cell, but again dummy bitlines and a new sense amplifier allow the read times to match the fast side of the cell regardless of the data being stored (as will be seen in section 3).

The speed enhanced (SE) cell in Fig. 2(b) dissipates higher leakage compared to the cell in Fig. 2(a) but it allows for read times that are virtually identical to that of the RV cell. Compared to the all RV cell the SE cell decreases leakage by more than 2X and 7X when holding a logical '0' and '1' respectively. The discharge time along BL is 61.2% longer compared to the RV cell, but dummy bitlines allow for quick sensing.

2.2 Supply Voltage Analysis

Leakage power is becoming increasingly important given the trend of decreasing the supply and threshold voltages in successive technologies [7]. We have tested our asymmetric cells with different supply voltages, and appropriately scaled threshold voltages, to measure leakage, discharge times, and cell flip times (the time required to flip the cell state). Fig. 3 shows the leakage while holding a '0' and '1' for all cells under different supply voltages. The figure shows that the leakage savings incurred by using the asymmetrical cells continues for lower supply voltages, and becomes more important as the leakage current rises exponentially with smaller supply voltages.

The bitline discharge times on the fast side of the cell and flip times for all cells are shown in Fig. 4(a) and (b) respectively. While the discharge time of the LE cell is slightly

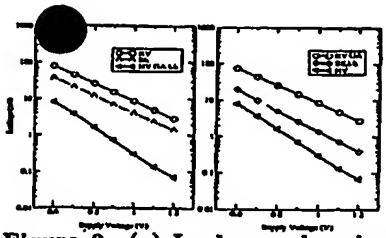

Figure 3: (a) Leakage when holding 0 (b) Leakage when holding 1

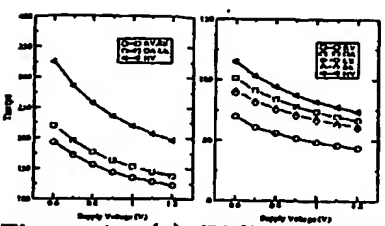

Figure 4: (a) Bitline discharge times (fast side) (b) Flip Times

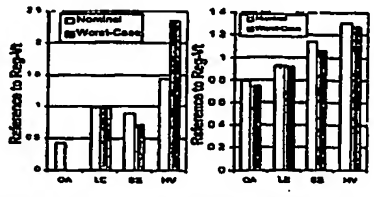

Figure 5: (a) Noise Margins (b) I_{trip}/I_{read} Stability

longer than that of the RV cell it is much shorter than that of the HV cell. The discharge time for the SE cell is virtually unchanged from that of the RV cell.

The cell flip times of the asymmetric cell all lie in between the cell flip times of the RV and HV cells, but, as seen in Fig. 4, are just a fraction of the discharge times.

2.3 Stability Analysis

Another major consideration with the cell design is its stability. There are two interrelated issues: read stability and noise margins [3][6]. Intuitively, read stability indicates how likely it is to invert the cell's stored value when accessing it, and was measured as the ratio of I_{trip}/I_{read} [3]. The static noise margin (SNM) of an SRAM cell is defined as the minimum dc noise voltage necessary to flip the state of the cell [8]. We have performed stability analysis on all the cells reported in this paper at a supply voltage of 1.2 V. Process variations were accounted for by performing the stability analysis under 59,049 combinations of different V_t and length for all six transistors in the cell. Fig. 5 shows the noise margins and stability of the cells normalized to those of the RV cell under nominal conditions and for the worst-case condition. While the OA cell fails under process variations, the LE and SE have comparable or better SNM and stability.

3. SENSE-AMPLIFIER

A conventional sense amplifier, shown in Fig. 6(a), is not suitable in our design due to the slow access time if the cell is storing a '0.' To obtain fast read times regardless of the data, a new sense amplifier was designed and is shown in Fig. 6(b). Compared to the conventional sense amp, the new sense amplifier has 4 extra transistors and an area increase of roughly 0.229 μm^2 or 14.4%. In addition, the sense amplifier uses a set of *dummy bitlines*, which are always fast (as fast as the fast side of the asymmetric cell), to trigger the reading of a logical '0' thus achieving fast access times when the slow bitline is discharging. Each pair of dummy bitlines are tied to the D and DB terminals of one column of dummy cells which all store a '1'. During every read operation one of the dummy cells will have its wordline asserted.

Sensing a '1' is as fast as a conventional sense amp since this is done by sensing a discharge of BLB due to the action of the fast side of the cell. Sensing a '0' is *initiated* at a later time than it would be in a conventional sense amp. This is done to allow sufficient time for the fast side to trigger the sense amp if it has to do so.

The sense amplifier operates as follows: Initially, the bitlines are precharged and all four amplifier inputs rise to VDD. If, during a read, BLB is being discharged (cell's fast side), then the differential pair composed of MN1 and MN2 causes increased current to pass through the left branch, thus increasing the voltage at node B and decreasing the

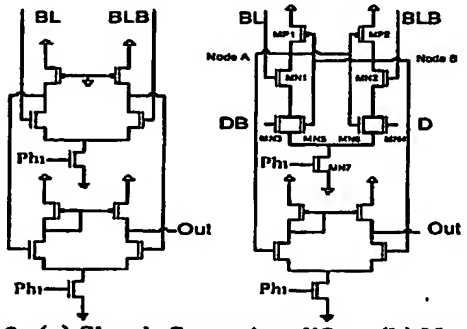

Figure 6: (a) Simple Sense Amplifier. (b) New Sense Amplifier

voltage at node A.

When BL is being discharged, then it does so at a slower rate since it is being discharged from the slow side of the asymmetric cell. To achieve fast sensing in this case also, the dummy bitlines, which are connected to the differential pair of MN3 and MN4, initiate the sensing of a logical '0.'

For this sensing scheme to achieve reliable results it must allow for adequate time for BLB to discharge before initiating a logical '0' read. This safety factor is achieved in two ways. First, the dummy bitlines are connected to all sense amps and therefore have a slightly higher capacitive load compared to real bitlines leading to a slower discharge on DB compared to BLB. The extra capacitive loading does not slow the sense time when BL is discharging because of the concerted effort between BL and DB to sense the same value. Second, the transistors connected to the bitlines are wider than the transistors connected to the dummy bitlines leading to a higher transconductance. This leads to a higher gain from the bitlines to the output than from the dummy bitlines. We have also performed sensitivity analysis of this sense amplifier, and it performs on par with the conventional sense amplifier.

4. SRAM

Using the above cells and the sense amplifier presented in section 3, a 32-Kbyte SRAM was designed and simulated to measure leakage, and read and write times. Each of the 128 SRAM sub-arrays contains 64 cells along each bitline, and 32 cells along each wordline. The SRAM was simulated at a temperature of 110°C with the RV, OA, LE, SE and HV cells. Furthermore the RV and HV cells were simulated with a conventional sense amp, and these results were used as a reference for our design.

Fig. 7 shows the total leakage within the SRAM attributable to the SRAM cells when the SRAM is either holding all '0's or all '1's. The leakage includes the leakage needed for the two sets of dummy cells. The leakage trends for the single cell in section 2 continue for the com-

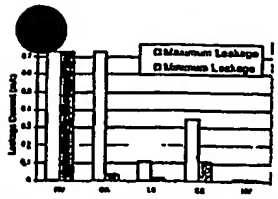

Figure 7: Max and Min Leakage Attrib. to Cells

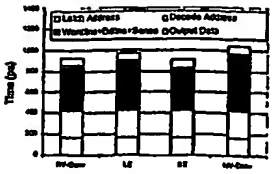

Figure 8: Breakdown of Memory Access Time

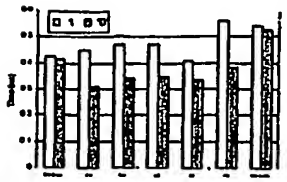

Figure 9: Sense times during a read cycle

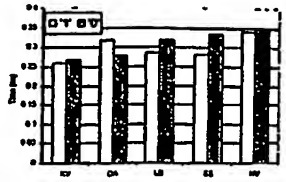

Figure 10: Write times for cells

plete SRAM where the LE and SE offer a reduction of 40X and 2X while storing a '0' and offer a reduction of about 7X when storing a '1.'

The total SRAM read access time includes four components: 1) input register propagation delay and hold times, 2) the address decoding delay, 3) the delay for wordline, bitline and sensing, and 4) the output register setup time. Our simulation results showing these components for the various SRAM arrays are shown in Fig. 8. Notice that only the 3rd component is affected by the cell design. Specifically, this time is the time period from when precharging is complete to when the sense amplifier has reached 90% of its swing.

Fig. 9(a) shows the sense times (the 3rd component of Fig. 8) for all cells. It can be seen that the worst-case sensing times are now on-par with the RV cell with a conventional sense amplifier. Compared with the RV cell with a conventional sense amp, the LE cell is 10% slower (although the total read time increases by about 5%, as seen in Fig. 8), but the SE cell is slightly faster (note this is not because the sense amplifier is quicker, but because the bitline discharge time for the SE cell is 50ps quicker than that of the RV cell, which is a byproduct of the asymmetry of the SE cell). The HV cell with a conventional sense amp would be 26% slower.

The write times for the different cells are shown in Fig. 10. The LE and SE cells show an increase of 19% and 25% respectively over the RV cell. The increase in write times is of minor importance since the write times are all shorter than the read times of the associated cells and therefore the speed of the SRAM is dependent on the read time.

5. ARCHITECTURAL ENHANCEMENTS

We investigated two cache organizations that use asymmetric cell designs: *statically biased* and *dynamic inversion*. In the *statically biased* cache, the cells are simply replaced with asymmetric ones. This cache is *statically biased* to dissipate low leakage power only when it stores the preferred bit value ('0'). What makes this cache successful is *typical* program behavior: as we show in [5], the SPEC2000 programs we studied exhibit a strong bias towards zero. The statically biased cache with the SE cells reduces leakage by 4.5X and 3.8X for an instruction and a data cache, respectively, compared to conventional symmetric-cell caches. The caches are 32Kbyte 4-way set associative caches. While programs with a higher fraction of '1's than '0's may exist, our SRAM would still dissipate much lower leakage power compared to the RV cell cache.

In *selective inversion*, the values stored within a block can be inverted at a byte granularity (other granularities are possible). In this design, if a byte contains five or more ones it is inverted prior to storing it in the cache. This cache needs an additional *inversion flag* cell per byte that holds information on which bytes were inverted. Inversion happens at write time. Since stores are typically buffered in a write buffer and are only sent to the data cache on commit, there is plenty of time to decide and apply inversion if necessary. Additional area, dynamic power and performance trade-offs apply to this design. An investigation of these issues is beyond the scope of this paper.

6. CONCLUSION

In this paper, we proposed a novel approach that combines both circuit- and architecture-level techniques. Our approach drastically reduces leakage power dissipation. The key observations behind our approach are that cache-resident memory values of ordinary programs exhibit a strong bias towards zero or one at the bit level.

We introduced a family of high-speed asymmetric dual-V_t SRAM cell designs that exploit this bit-level bias to reduce leakage power while maintaining high performance. The *speed enhanced* cell reduces leakage power by at least 2X and by 6X in the preferred state. It is as fast as the conventional, regular-V_t SRAM cell. By comparison, the *leakage enhanced* cell reduces leakage by at least 6X and by about 40X in the preferred state. Its sense time is 10% higher than the speed enhanced and the regular-V_t cells (total read time is only 5% higher).

7. REFERENCES

[1] S. Borkar, "Design challenges of technology scaling," *IEEE MICRO*,vol.19, no.4, pp. 23–29, July–Aug. 1999.

[2] T. Kam et. al., "EDA challenges facing future microprocessor design," in *IEEE Transactions on Computer-Aided Design*, vol. 19, no. 12, Dec. 2000.

[3] F. Hamzaoglu et. al., "Dual-Vt SRAM cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13um technology generation," in *Proc. 2000 Intl. Symp. on Low Power Electronics and Design*, July 2000.

[4] John L. Hennessy and David A. Patterson, Computer Architecture: A Quantitative Approach (2nd edition), Morgan Kaufman, 1996.

[5] N. Azizi, A. Moshovos, F. N. Najm, B. Falsafi, "Asymmetric-cell caches: exploiting bit value biases to reduce leakage power in deep-submicron, high-performance caches," *ECE Computer Group Technical Report TR-01-01-02*, Univ. of Toronto.

[6] E. Seevinck Sr et. al., "Static-noise margin analysis of MOS SRAM cells," *IEEE Journal of Solid-State Circuits*, vol. 22, pp. 748–754, Oct. 1987.

[7] V. De, and S. Borkar, "Technology and Design Challenges for Low Power and High Performance Microprocessors," in*Proc. 1999 Intl. Symp. on Low Power Electronics and Design*, 1999.

[8] A. Bhavnagarwala et. al., "The Impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability," in *IEEE J. of Solid-State Circuits*, vol. 36, Apr. 2001.

Low-Leakage Asymmetric-Cell SRAM†

Navid Azizi
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
nazizi@eecg.toronto.edu

Farid N. Najm
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
f.najm@utoronto.ca

Andreas Moshovos
ECE Department
University of Toronto
Toronto, Ontario, Canada
M5S 3G4
moshovos@eecg.toronto.edu

Abstract— **We introduce a novel family of asymmetric dual-V_t SRAM cell designs that reduce leakage power in caches while maintaining low access latency. Our designs exploit the strong bias towards zero at the bit level exhibited by the memory value stream of ordinary programs. Compared to conventional symmetric high-performance cells, our cells offer significant leakage reduction in the zero state and in some cases also in the one state albeit to a lesser extent. A novel sense-amplifier, in combination with dummy bitlines, allows for read times to be on par with conventional symmetric cells. With one cell design, leakage is reduced by 7X (in the zero state) with no performance degradation, but with a stability degradation of 6%. Another cell design reduces leakage by 2.3X (in the zero state) with no performance or stability loss. An alternative cell design reduces leakage by 58X (in the zero state) with a performance degradation of 1% and an area increase of 2.4% and no stability degradation.**

I. Introduction

As a result of technology trends, leakage (static) power dissipation has emerged as a first-class design consideration in high-performance processor design. Historically, architectural innovations for improving performance relied on exploiting ever larger numbers of transistors operating at higher frequencies. To keep the resulting switching power dissipation at bay, successive technology generations have relied on reducing the supply voltage. In order to maintain performance, however, this has required a corresponding reduction in the transistor threshold voltage. Since the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) sub-threshold leakage current increases exponentially with a reduced threshold voltage, leakage power dissipation has grown to be a significant fraction of overall chip power dissipation in modern, deep-submicron ($< 0.18\mu$m) processes. Moreover, it is expected to grow by a factor of five every chip generation [1]. For processors it is estimated that in 0.10μm technology, leakage power will account for about 50% of the total chip power [2].

Since leakage power is proportional to the number of transistors, and given the projected large memory content of future System-on-Chip (SOC) devices, it becomes important to focus on Static Random Access Memory (SRAM) structures such as caches, which comprise the vast majority of on-chip transistors. Existing circuit-level leakage reduction techniques are oblivious to program behavior and trade off performance for reduced leakage where possible, e.g., [3]. Combined circuit- and architecture-level techniques reduce leakage for those parts of the on-chip caches that remain unused for long periods of time (thousands of cycles) [4][5][6]. The mechanisms that identify which cache parts will be unused and that enable leakage reduction incur considerable power and performance overheads that have to be amortized over long periods of time. These methods are not effective when most of the cache is actively used.

We present a family of novel *asymmetric* SRAM cell designs that lead to new cache designs which we refer to as the Asymmetric-Cell Caches (ACC). ACCs offer drastically reduced leakage power compared to conventional caches even when there are few parts of the cache that are left unused. ACCs exploit the fact that in ordinary programs most of the *bits* in caches are *zeroes* for both the data and instruction streams. It has been shown that this behavior persists for a variety of programs under different assumptions about cache sizes, organization and instruction set architectures, even when assuming perfect knowledge of which cache parts will be left unused for long periods of time [7].

Traditional SRAM cells are symmetrically composed of transistors with identical leakage and threshold characteristics. Our *asymmetric* SRAM cell designs offer low leakage with little or no impact on latency. In our asymmetric SRAM cells, selected transistors are "weakened" to reduce leakage when the cell is storing a zero (the common case). In this work, we achieve the weakening by using higher-V_t transistors, however, this may also be possible by appropriate transistor sizing. We evaluate our designs by simulation, based on a commercial 0.13μm, 1.2V CMOS technology. The six best designs offer different performance/leakage/stability characteristics. With one cell, leakage is reduced by 7X (in the zero state) with no performance degradation. An alternative cell design reduces leakage by 70X (in the zero state) with a read time degradation of 5%. These cells have slightly lower stability; four other cells with improved stability are also presented with leakage reductions of up to 58X.

We make the following contributions: (1) We introduce a novel family of asymmetric SRAM cells. No previous work on designing asymmetric SRAM cells exists. (2) We introduce a novel sense amp design that exploits the asymmetric nature of our cells to offer cell read times that are on par with conventional symmetric SRAM cells. (3) We evaluate a cache de-

†This project was supported in part by the Semiconductor Research Corporation (SRC 2001-HJ-901).

sign t[illegible]s based on ACCs and demonstrate that compared to a conventional cache, it offers drastic leakage reduction while maintaining high performance and comparable noise margins and stability.

The rest of this paper is organized as follows: In section 2, we present our asymmetric cell family. In section 3, we present the sense amplifier. In section 4, we present the simulation results of an SRAM using the different asymmetric cells. Section 5 includes a discussion on architectural level techniques to leverage the asymmetric nature of the cells. Finally, we conclude the paper in section 6.

II. Asymmetric SRAM Cells

Early work done to reduce the power consumption of SRAM's consisted of reducing the dynamic power dissipation through changes in the peripheral circuitry. Due to the large number of transistors contained in SRAM arrays, the static power dissipation within the array has become a large fraction of the total power dissipation. Ideally, an SRAM cell should be fast and should dissipate low leakage power. This is increasingly at odds with the fundamental technology trade off between transistor speed and leakage. Conventional high-performance SRAM cells use a symmetric configuration of six transistors with identical threshold voltages. One can reduce leakage by using higher-V_t transistors, but unfortunately using an all-high-V_t transistor cell degrades performance by an unacceptable margin. Our asymmetric SRAM cells reduce leakage while maintaining high performance based on the following approach: select a preferred state and weaken only those transistors necessary to drastically reduce leakage when the cell is in that state. These cells exhibit asymmetric leakage and access behavior. Fortunately, their asymmetric access behavior can be exploited to maintain high performance while reducing leakage.

A. Technology

All results reported in this paper are HSPICE simulation results produced at 110°C using SPICE models of a commercial 0.13μm, 1.2V CMOS technology. Furthermore, throughout this paper, the following convention will be used. A High-V_t (HV) transistor is obtained from the basic 0.13μm, 1.2V, transistor (referred to herein as the Regular-V_t (RV) transistor) by artificially increasing the V_t by 0.2V using the HSPICE in-line parameter DELVTO. This value of 0.2V was chosen because it leads to a difference of about 10X between the leakage currents of HV and RV transistors, which is typical of dual-V_t technology. Finally, the sizes of the transistors comprising the basic SRAM cell, which forms the starting point for our design variations below, are part of the technology specification for the 0.13μm process. As such, these sizes cannot be disclosed.

B. Nine Asymmetric Cells

As shown in Fig. 1, an SRAM cell comprises two inverters, (P2, N2) and (P1,N1), and two pass transistors, N3 and N4. In the inactive state, the wordline (WL) is held low so that the two pass transistors are off isolating the cell from bitline (BL) and bitline-bar (BLB). At this stage the bitlines are also typically

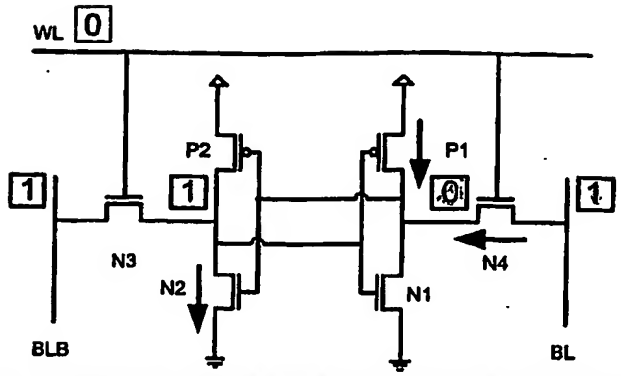

Fig. 1. Transistors that dissipate leakage when an SRAM cell is holding a '0'

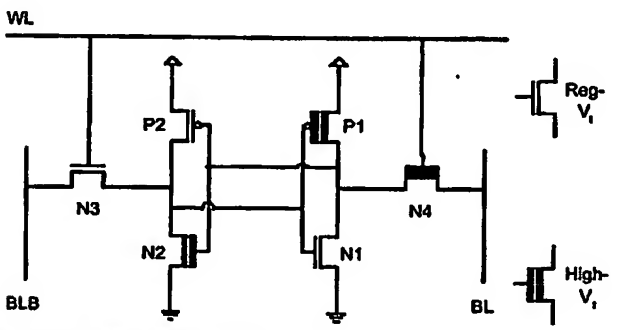

Fig. 2. Basic Asymmetric SRAM cell

charged at V_{DD} (e.g., logic '1'). Cells spend most of their time in the inactive state. In this state, most of the leakage is dissipated by the transistors that are *i)* off and that *ii)* have a voltage differential across their drain and source. The value stored in the cell (i.e., the cell state) determines which transistors these are. When the cell is storing a '0', as in Fig. 1, the leaky transistors are P1, N4 and N2. If the cell was storing a '1' then transistors P2, N1 and N3 would dissipate leakage power.

A simple technique for reducing leakage power would be to replace all transistors with high-V_t ones, but this unacceptably degrades the bitlines discharge times by 61.6% *. Since ordinary programs exhibit a strong bias in cache-resident bit values [8], another possibility to reduce leakage power, but at the same time keep read access times short, is to choose a preferred stored value and to only replace those transistors that contribute to the leakage power in this state with HV transistors, as seen in Fig. 2. This Basic Asymmetric (BA) cell was simulated and it exhibits the same leakage as the RV cell when holding a logic '1', but its leakage is reduced by 70X when holding a logic '0.'

The read access time of the BA cell is, however, degraded. Due to N2's and N4's higher threshold voltage, the bitline discharge takes longer. The discharge times for BLB and BL are 12.2% and 46.4% longer than the discharge time for the RV cell, respectively. Read times can be made to match the faster read time by using a set of dummy bitlines and a novel sense amplifier, as discussed in Section III.

Since p-Channel Metal Oxide Semiconductor (PMOS) tran-

*Discharge time is defined as the time from when the wordline is raised to when one of the bitlines reduces to 90% of its precharge value. 90% was chosen due to it being a appropriate differential signal for sense amplifiers to trigger.

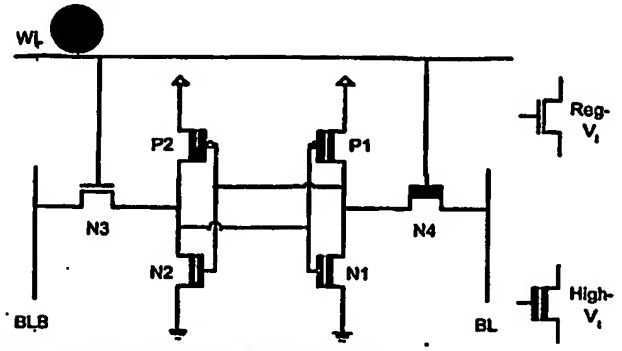

Fig. 3. Leakage Improved 3 Cell (the LE Cell)

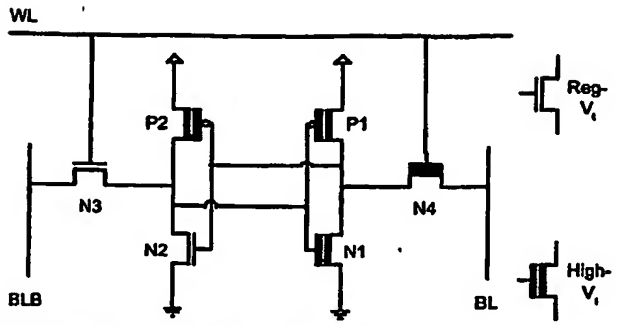

Fig. 4. Speed Improved 3 Cell (the SE Cell)

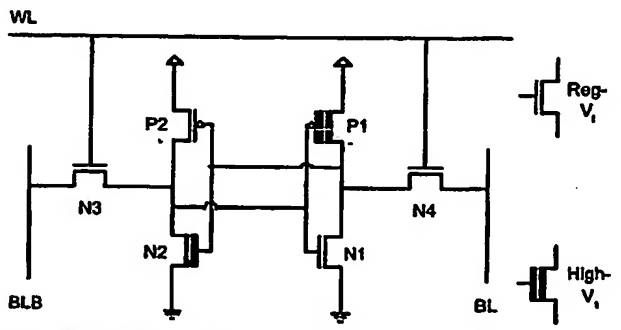

Fig. 5. Special Precharge Cell

sistors have very little effect on the cell's read access time (the role of pulling down the bitlines is played by the two n-Channel Metal Oxide Semiconductor (NMOS) transistors on the side of the cell storing the '0'), a better asymmetric cell consists of the BA cell with P2 also set to high-V_t. This cell, referred to as Leakage Improved 2 (LI2), has the advantage of partially reduced leakage in the high leakage state. When the cell is holding a logic '1' its leakage is reduced by 1.6X relative to the RV cell, and when holding a logic '0' its leakage is reduced by 70X. The discharge times for BLB and BL are 12.2% and 46.4% longer than the discharge times for the RV cell, respectively, the same as the BA cell's discharge times. One further improvement is possible because, due to the sense amplifier (described below) which matches the read time on the slow side of the cell to the fast side, there is no need for N1 to be low-V_t. This leads to the cell in Fig. 3, referred to as Leakage Improved 3 (LI3). This cell further reduces leakage in the high leakage state, so that its leakage relative to the RV cell is reduced by 7X in the '1' state and by 70X in the '0' state. The BL discharge time is now 61.6% longer than the discharge time for the RV cell, but that is of minor importance due to the novel sense amplifier design, as we will see below.

The two asymmetric cells, LI2 and LI3, take the BA cell and improve its leakage performance while not affecting its read access time. Another design front is to take the BA cell and try to improve its read access time, while keeping some of the leakage benefits found in the BA cell.

To eliminate the speed penalty incurred in the BA cell due to both pull-down paths having one high-V_t transistor, both N2 and N3 are made low-V_t. This cell, Speed Improved 1 (SI1), now has discharge times for BLB and BL which are 0% and 46.7% respectively longer than the RV cell. Thus one side of the cell is just as fast as the RV cell. However, this cell suffers from higher leakage than the BA cell, with a leakage reduction of 2X relative to RV when holding a '0', and no leakage reduction when holding a '1'.

The same transformations performed on BA to improve its leakage performance can also be performed on the SI1 cell. First P2, is made high-V_t, and then N1 is also made high-V_t. These two new cells are named Speed Improved 2 (SI2) and Speed Improved 3 (SI3), respectively. SI3 is shown in Fig. 4. SI2 has leakage reductions of 2X and 1.6X when storing a '0' and '1', respectively, while SI3 has leakage reductions of 2X and 7X. † These two cells have no read access time degradation compared to the RV cell along BLB, but have a 46.5% and 61.6% degradation along BL respectively. Once again, the degradation along BL is of minor importance due to the novel sense amplifier.

One would like to get the very low leakage of the LI2 and LI3 cells and a very small read access delay. A final asymmetric cell can meet these objectives, but it requires a different read operation. In the steady state, instead of keeping BL precharged to V_{DD}, keep it at ground. Now, N4 can be kept low-V_t for the preferred '0' state. This Special Precharge (SP) cell is shown in Fig. 5. Before a read, BL may have to be raised to '1', or a new sensing scheme may have to be developed, which may be power hungry. This cell requires changes to the peripheral circuits of the SRAM array, and further work is required to develop this concept. Nevertheless, the results for this cell are presented for completeness: leakage is reduced by 83.3X in the '0' state, while the '1' state shows no leakage reduction. Bitline discharge times are degraded by 12.2% and 0%.

A summary of the leakage reduction while holding a '0' and '1' can be seen in Table I and Fig. 6. The bitline discharge times are summarized in Table II, which shows the distinction between the Leakage Improved (LI) and Speed Improved (SI) cells. All LI cells show a near 12% increase in bitline discharge times, while the SI cells show no increase in bitline discharge times. Furthermore, Fig. 6 shows that the BA and LI2 cells show no advantage to LI3, since the LI3 cell has the same speed performance as the BA and LI2 cells but with better leakage

†Note that SI3 reverses the preferred leakage state to the state when the cell is holding a '1'. All further references to this cell will have the '1' state as the preferred state so that the cell language remains in conformity with all other cells, but it should be noted that in practice the cell bitlines can be flipped to allow for '0' to be the preferred state without affecting any of the performance or stability results shown here.

TABLE I
SUMMARY OF LEAKAGE REDUCTION FOR ALL ASYMMETRIC CELLS, RELATIVE TO THE RV CELL

Asymmetric Cell	Leakage Reduction storing a '0' (times)	Leakage Reduction storing a '1' (times)
BA	69.50X	1.00X
LI2	69.50X	1.61X
LI3	69.50X	6.96X
SI1	2.04X	1.00X
SI2	2.04X	1.61X
SI3	2.04X	6.96X
SP	83.33X	1.00X
HV	69.50X	69.50X

TABLE II
SUMMARY OF BITLINE DISCHARGE TIMES FOR ALL ASYMMETRIC CELLS, RELATIVE TO THE RV CELL

Asymmetric Cell	% Increase of BLB discharge time	% Increase of BL discharge time
BA	12.25%	46.73%
LI2	12.25%	46.50%
LI3	12.09%	61.64%
SI1	0.00%	46.73%
SI2	0.00%	46.51%
SI3	0.00%	61.69%
SP	12.26%	0.00%
HV	61.64%	61.64%

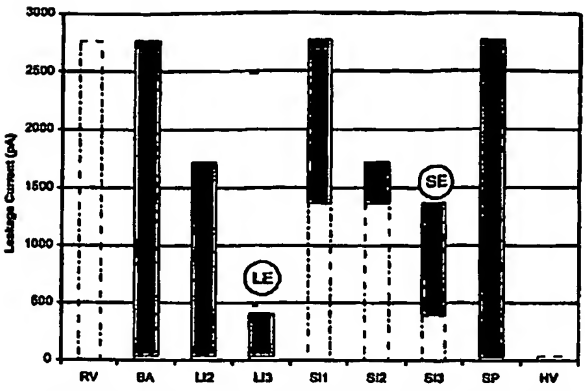

Fig. 6. Graphical Representation of Asymmetric Leakage Characteristics of all cells

TABLE III
SUMMARY OF WRITE TIMES FOR ALL ASYMMETRIC CELLS

Asymmetric Cell	Percent Increase over RV cell
BA	51.9%
LI2	41.2%
LI3	36.0%
SI1	56.1%
SI2	45.5%
SI3	40.2%
SP	11.5%
HV	69.4%

performance. The LI3 cell will be referred to henceforth as the Leakage Enhanced (LE) cell. Also, SI3 is clearly the best design from within the SI cells. The SI3 cell will be referred to henceforth as the Speed Enhanced (SE) cell.

Until now, only the bitline discharge times of the different cells have been compared, and write times have been ignored. The write times of the cells are less important because stronger write drivers can be designed to drive the bitlines, and write drivers are a small portion of the total SRAM. The write times of the asymmetric cells all lie within the write times of the RV cell and the HV cell. The precise numbers can be seen in Table III. The LE and SE cells show the smallest increase in their respective groupings.

Since the LE and SE are the two best designs from the two sets of asymmetric cells, only these two cells, and variations on them, will be further discussed in the following.

C. Stability

Another major consideration with the cell design is its stability. There are two interrelated issues: read stability and noise margins [3][9]. Intuitively, read stability indicates how likely it is to invert the cell's stored value when accessing it, and was computed as the ratio of I_{trip}/I_{read}, where I_{trip} is the current through the pull-down NMOS when the state of the cell is being reversed by injecting an external current I_{test} and where I_{read} is the maximum current through the pass transistor during a read [3]. The static noise margin (SNM) of an SRAM cell is defined as the minimum dc noise voltage necessary to flip the state of the cell [10]. In our case, the stability of all cells was measured by simulation (HSPICE) via both the Static Noise Margin (SNM) and the I_{trip}/I_{read} methods. Under both stability tests, the stability was first measured under nominal conditions, assuming no process variations.

Then, to measure stability under process variations, two sets of tests were performed. First, the SNM and I_{trip}/I_{read} tests were performed on 59,049 combinations of different V_t and length variations for all six transistors in the cell. The combinations included modifying by $\{-3\sigma, 0, +3\sigma\}$ the NMOS transistors' V_t and length values and the PMOS transistors' V_t value thus giving $3^8 = 59,049$ combinations. The worst-case value for various cells was found, and compared to the worst-case value obtained for the RV cell.

Second, Monte-Carlo analysis was performed to obtain a distribution for the SNM and I_{trip}/I_{read}. For each cell, 500 scenarios for V_t and length were randomly generated, consistent with their joint distributions, and simulated. The mean of the distribution was estimated using the unbiased estimator in (1), and the variance was estimated by using the unbiased estimator in (2). Furthermore, the Normal Scores Method was used to graphically determine the distribution type [11]. Given the distribution type, mean, and variance, the probability of failure for various cells was then computed.

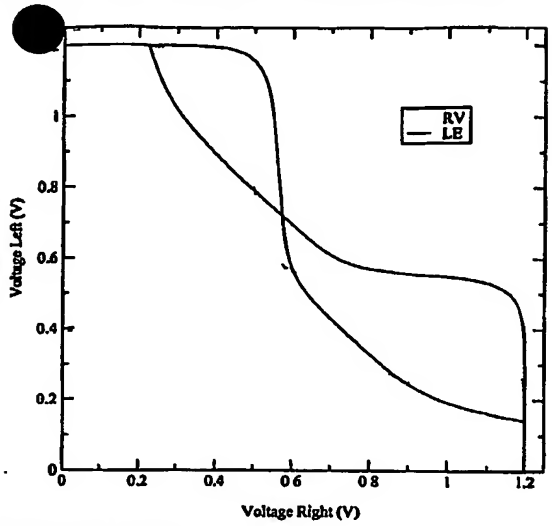

Fig. 7. SNM of the LE cell

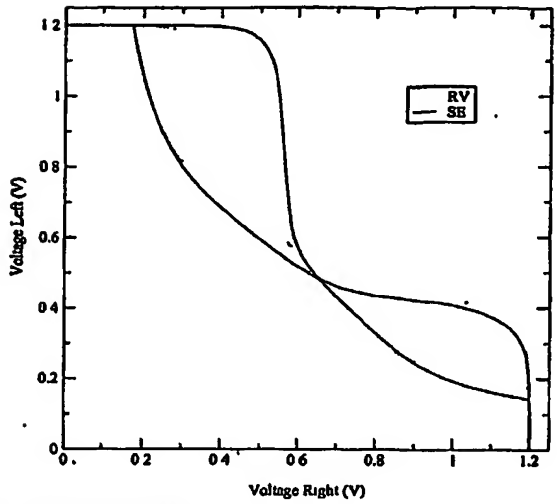

Fig. 8. SNM of the SE cell

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (1)$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad (2)$$

1) Static Noise Margin: The SNM of the LE and SE cells were computed through simulation. The SNM of the RV cell was also computed to be used as a reference. Under nominal conditions, the SNM of the LE and SE cells were $0.246V$ and $0.221V$, respectively, while the SNM of the RV cell was $0.250V$. Thus, the LE and SE show a decrease in SNM of 1.6% and 11.7%. One would expect that by using HV transistors in the design the SNM of the cells would increase, but the asymmetry of the cells skews the lobes of the butterfly curve and decreases the SNM, as will be explained below.

First, let us examine the SNM of the cells when the wordline is not active. During this state, the SRAM cell is not as vulnerable as when it is being read, but a study of this case helps to understand the decrease in the SNM when the cell is being read. When the wordline is off, the only transistors that affect the SNM are the four transistors comprising the back-to-back inverters.

Since the four internal transistors of the LE cell are all high-V_t, the cell has equal low and high noise margins of $0.685V$, a 22.6% increase over the standby SNM of the RV cell, $0.559V$. However, when the SNM of the cell is being measured during a read, as seen in Fig. 7, the cell has high SNM in one state, $0.363V$, and low SNM in the other, $0.246V$. The asymmetry in the LE butterfly curve is due to the mismatch between the strength of the pass-gate (N3) and pull-down (N2) transistors. During a read, the N3 pass transistor, due to it being low-V_t, has a higher conductivity than N2 and raises the voltage at the storage node to a higher voltage than if the two NMOS were of equal strength.

For the SE cell, the internal inverter pair are not identical. Thus the standby (i.e., with the wordline off) SNM of the cell has asymmetric lobes with noise margins of $0.535V$ and $0.727V$, in the worst-case a 4.2% decrease in noise margin compared to the RV cell. The source of this mismatch is the V_t difference between N1 and N2, which causes one of the transfer characteristics to commence its transition in the SNM plot from '0' to '1' later than normal. During a read, the mismatch between the size of the lobes becomes exaggerated because it is as if a constant is subtracted from the noise margin on each side of the cell since each side of the cell has equal strength pass transistors and pull-down transistors. While being read, the SE cell has low and high noise margins of $0.222V$ and $0.365V$ respectively. The SNM plot of the SE cell is shown in Fig. 8.

TABLE IV
WORST-CASE SNM

Cell	Worst-Case SNM(V)	% Increase over RV cell
RV	0.091	–
LE	0.088	-3.73%
SE	0.065	-28.79%

As explained in Section II-C, process variations were analyzed by two methods. First, by sweeping over 59,049 cases, the worst-case SNM was found for each cell and is summarized in Table IV.

The asymmetric cells stability performance degrades compared to that of the RV cell. Since process variations induce an asymmetry in the butterfly curve, the original asymmetry inherent in the butterfly curves for the LE and SE allows one lobe of the butterfly curve to become pinched off even further and lose stability. For the LE cell the butterfly curve becomes pinched off when N3 becomes stronger than N2 and P1 increases in strength, while N1 does not. Fig. 9 shows the effect graphically. The worst case for the SE cell occurs at a different process corner. The butterfly curve becomes pinched off when P2 decreases in strength and N2 increases in strength, and N4 gets stronger than N1, as shown in Fig. 10

Monte-Carlo Analysis was also performed on the RV, LE and SE cells. Table V summarizes the mean and standard deviation of the SNM. Furthermore, the Normal Scores method showed

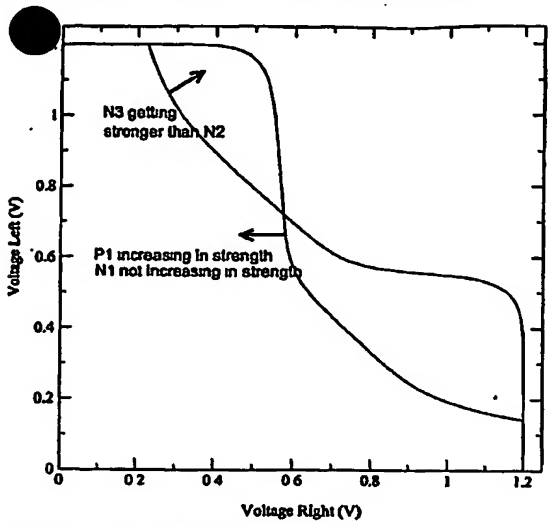

Fig. 9. SNM of LE cell pinching off

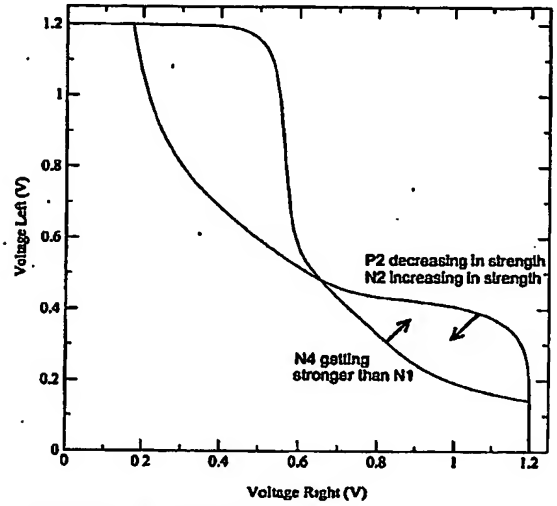

Fig. 10. SNM of SE cell pinching off

that the distributions for all cells were Gaussian. Due to their very small standard deviation, the SNM of all cells remains very close to their respective mean [‡]. Thus the mean of the SNM becomes a very important measure, and is a better reflection of the stability than the nominal or worst-case SNM. Using the mean as a measure of stability, the LE has a 7% increase in SNM and the SE has a 5.8% decrease.

2) I_{trip}/I_{read}: Using the SNM as a measure of stability showed that the LE cell was comparable to the RV cell while the SE cell showed a marginal decrease in stability. When I_{trip}/I_{read} is computed by simulation, it is seen that the SE outperforms the RV cell and the LE suffers. Table VI shows the results.

The LE cell has a lower I_{trip}/I_{read} value due to the V_t mismatch between the pass transistor and pull-down transistor on one side of the cell. The I_{trip} values from both sides of the cell show a drop compared to the I_{trip} value from the RV cell due to both pull-down transistors becoming high-V_t. However, with N3 remaining low-V_t, I_{read} on the fast side of the cell does not suffer the same drop, and I_{trip}/I_{read} falls compared to that of the RV cell.

[‡]For example, by using the inverse *erfc* function, it was found that the probability for the SE cell to have an SNM of 0.212 (a mere 2% drop of the SNM) was only 5.421×10^{-18}.

TABLE V
MEAN AND σ DURING MONTE-CARLO ANALYSIS FOR SNM

Cell	Mean(V)	Standard Deviation(V)
RV	0.231	0.0182
LE	0.247	0.0244
SE	0.218	0.0249

The SE, due to it having the same strength pull-down and pass transistors on each side of the cell, does not experience the same problem as the LE cell. On the slow side of the cell, both I_{trip} and I_{read} fall compared to the RV cell, but I_{read} falls by a larger amount thus increasing the I_{trip}/I_{read}. On the fast side of the cell, I_{read} does not change compared to the RV cell, but I_{trip} increases slightly. In the RV cell, the reduction in voltage (due to leakage) at the stored '1' node degrades the current sinking capacity of the pull-down NMOS. In the SE cell, because of the high-V_t transistors on the '1' side of the cell there is no degradation in the current sinking capacity of the pull-down transistor and thus I_{trip} increases leading to a larger I_{trip}/I_{read}.

A total of 59,049 different corner cases of process variations were simulated and the worst-case I_{trip}/I_{read} was noted in each cell, and is summarized in Table VII. The LE cell and the RV cell achieve their worst-case I_{trip}/I_{read} for the same process corner: when the difference in strength between N2 and N3 is amplified with N2 becoming weaker, and N3 becoming stronger. The SE cell, however, suffers its worst-case I_{trip}/I_{read} when N4 becomes stronger than N1.

The Monte-Carlo analysis show that I_{trip}/I_{read} is also Gaussian from the linear plots obtained from the Normal Scores Method. Table VIII shows the mean and standard deviation of the three cells. Notice, once again the standard deviation is very small, and thus most cells will be very near the mean where the LE shows a 4.35% decrease and the SE cell shows a 14.84% increase in I_{trip}/I_{read}.

D. Improved-Stability Cells through Threshold Voltage

As seen from the previous section, the SE and LE cells have either a lower stability in the SNM test or the I_{trip}/I_{read} test. In many cases, the stability of the cell is a critical factor to obtain a desired yield and to lower the cost of the chip. In that regard, two derivative cells, one from the LE cell and one from the SE, have been developed that improve upon their SNM, but do not decrease the leakage as much as the SE and LE cells. The two new cells are named Stability-Leakage Enhanced (SLE) and Stability-Speed Enhanced (SSE). [§]

One way to improve the SNM of the cells under process variations is to try to make the size of the lobes of the butterfly curve

[§]The I_{trip}/I_{read} of the cells is not improved since the only method to improve the LE cell's value considerably is to make the pull-down NMOS on the fast side of the cell low-V_t which would make it the same as the SE cell. The SE cell already has a better I_{trip}/I_{read} than RV under nominal, worst-case, and mean conditions.

TABLE VI
NOMINAL I_{trip}/I_{read}

Cell	Nominal(V)	% Increase
RV	2.26	–
LE	2.10	-7.31%
SE	2.60	14.86%

TABLE VII
WORST-CASE I_{trip}/I_{read}

Cell	Worst-Case(V)	% Increase
RV	1.71	–
LE	1.58	-7.81%
SE	1.82	6.12%

TABLE VIII
MONTE-CARLO ANALYSIS FOR I_{trip}/I_{read}

Cell	Mean	σ
RV	2.20	0.0765
LE	2.10	0.0898
SE	2.53	0.1100

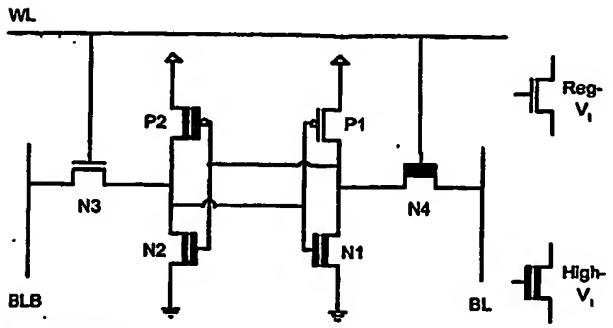

Fig. 11. SLE cell

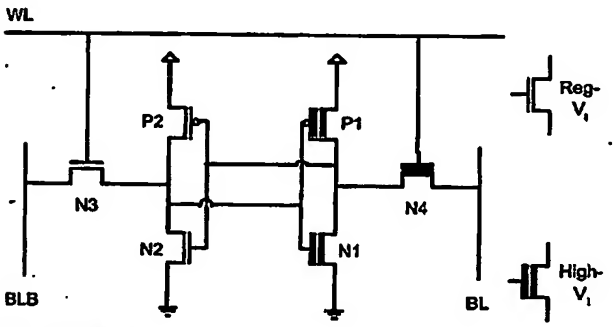

Fig. 12. SSE cell

symmetric. For the LE cell the lobes can be made more symmetric by making N2 low-V_t, but this new cell would just be the SE cell. Another option is to make P1 low-V_t. This change, seen in Fig. 11, has the opposite effect of the lower arrow in Fig. 9, and makes the lobes of the butterfly curve more symmetric. The SNMs are now $0.360V$ and $0.283V$ instead of $0.363V$ and $0.246V$. The change in SNM can be seen in Fig. 13. To make the SE cell's SNM plot more symmetric, P2 can be made low-V_t to have the opposite effect as the top arrow in Fig. 10. By doing this, the SNM plot shown in Fig. 14 has SNMs of $0.256V$ and $0.362V$ instead of $0.222V$ and $0.366V$. The SSE cell is shown in Fig. 12.

For these stability-improved cells, all the previous tests for leakage, performance, and stability can be performed to compare them to the cells they were derived from, as well as to the RV cell.

1) Leakage: The leakage performance of the stability-improved cells falls off, as is expected due to one transistor in the LE and SE being re-converted to a low-V_t transistor. For the SLE cell, the leakage reduction when holding a '1' remains unchanged at a 6.96X reduction relative to RV, but the leakage reduction when holding a '0' changes from 69.5X to 2.5X.

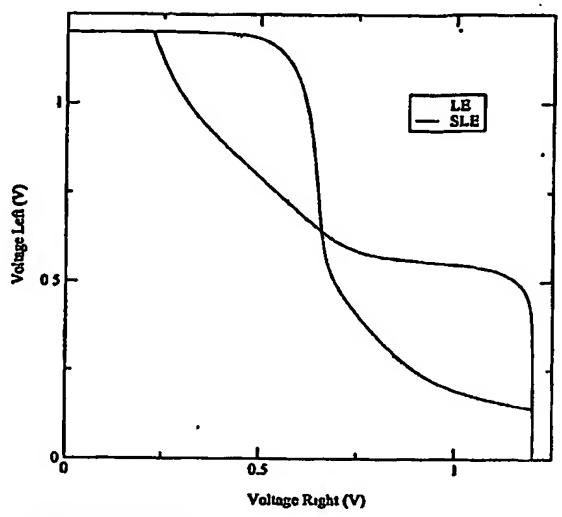

Fig. 13. SNM for SLE cell

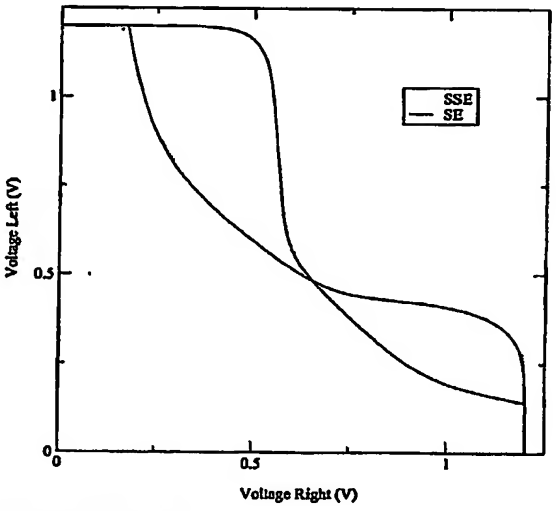

Fig. 14. SNM for SSE cell

For the SSE cell, when it is holding a '0' the leakage reduction stays at 2.04X, but when it is holding a '1' the leakage reduction changes from 6.96X to 1.91X.

2) Performance: Since the PMOS transistors do not play a large role in discharging the bitlines, it would be expected that the discharge time for the stability-improved cells to be very close to the cells they derived from. Through simulation, it is seen that the discharge times along BL and BLB remain almost constant. As for the write times, SLE's write time decreases to a 33.15% increase over RV's write time from LE's 35.95%

TABLE IX
% IMPROVEMENT OVER RV CELL FOR IMPROVED-STABILITY CELLS UNDER SNM TEST

Cell	Worst-Case SNM % improvement	Mean SNM % improvement
SLE	35.41%	22.63%
SSE	7.80%	9.29%

TABLE X
% IMPROVEMENT OVER RV CELL FOR INCREASED-STABILITY CELLS UNDER I_{trip}/I_{read} TEST

Cell	Worst-Case I_{trip}/I_{read} % improvement	Mean I_{trip}/I_{read} % improvement
SLE	-10.50%	-6.78%
SSE	4.06%	13.43%

increase. The SSE write time jumps to a 49.22% increase over RV's write times.

3) Stability: The stability analysis has also been performed on the derivative cells for both the SNM and I_{trip}/I_{read}. Both derivative cells perform better than the RV cell in the worst case, and under Monte-Carlo analysis. The results are shown in Table IX.

Under the I_{trip}/I_{read} method, there is very little change, because I_{trip}/I_{read} depends strongly on the NMOS transistors, which have not been changed, but the stability-improved cells perform slightly worse than the cells from which they were derived. The results are summarized in Table X.

E. Improved-Stability Cells through Transistor Sizing

From the previous section, it can be seen that when stability is recovered through a change in threshold voltage of the PMOS transistors, a large portion of the leakage benefits of the asymmetric cells are lost. Furthermore, the low I_{trip}/I_{read} of the LE cell could not be improved by threshold voltage assignment.

Another way of improving stability is to resize some of the transistors to reclaim the conductance lost due to the high-V_t assignment. This change does not have a large effect on the leakage characteristics because leakage increases exponentially with reduced threshold voltages, but increases only linearly with transistor size. Moreover, the low I_{trip}/I_{read} of the LE cell can be improved by transistor resizing.

The lobes of the SNM plot for the SE cell can be made more symmetric by making N1 wider. In our case, we increased the width of this transistor by 26%, leading to a new cell which we refer to as Resized Speed Enhanced (RSE). The SNM for the RSE cell is comparable to that of the RV cell and the change in N1's size leads to an increase of only 2.9% in cell area. The change in the SNM plot can be seen in Fig. 15, where the margins are now $0.253V$ and $0.347V$ instead of $0.222V$ and $0.366V$. The RSE cell's nominal value for I_{trip}/I_{read} does not change much compared to the nominal value for the SE cell. On the slow side of the cell, which had the higher I_{trip}/I_{read} value

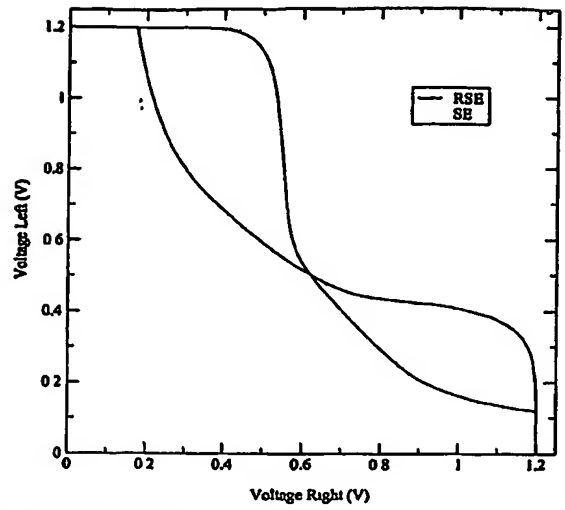

Fig. 15. SNM for RSE cell

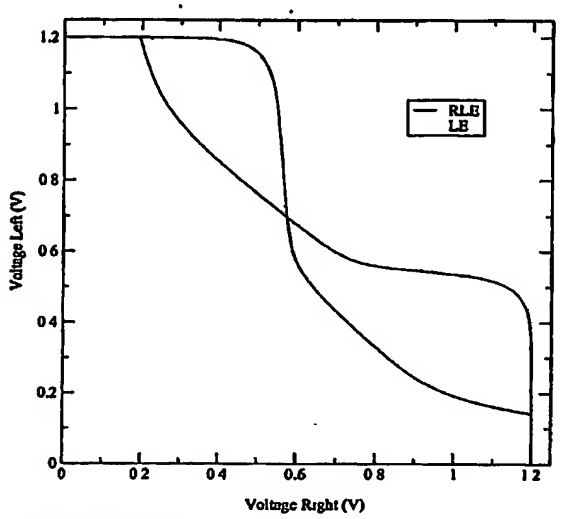

Fig. 16. SNM for RLE cell

for the SE cell, the increase in N1's size allows for I_{trip} to become larger and increases the I_{trip}/I_{read} value. The fast side of the cell however, which has the limiting I_{trip}/I_{read} value, has a reduced I_{trip} that reduces the final value of I_{trip}/I_{read} to 2.53. The reduction in I_{trip} is due to the '1' storage node having a slightly lower voltage due to the increased leakage through N1. Nevertheless, the RSE cell's I_{trip}/I_{read} value is still 11.8% better than that of the RV cell.

For the LE cell, increasing the width of N2 allows the conductance of N2 to approach that of N3, which leads to an increase in I_{trip}, thus increasing I_{trip}/I_{read}. By increasing N2's width by 22%, (leading to an only 2.4% increase in cell area) the I_{trip}/I_{read} value of the the new Resized Leakage Enhanced (RLE) cell was made to be 2.28, which is comparable to the I_{trip}/I_{read} value of 2.26 of the RV cell. The increase in N2's width also increases the SNM of the RLE cell where the margins are now $0.349V$ and $0.280V$ instead of $0.363V$ and $0.246V$. The SNM plot for this cell can be seen in Fig. 16.

For these resized cells, all the previous tests for leakage, performance, and stability were performed to compare them to the

TABLE XI
% IMPROVEMENT OVER RV CELL FOR RESIZED-STABILITY CELLS UNDER SNM TEST

Cell	Worst-Case SNM % improvement	Mean SNM % improvement
RLE	36.65%	21.4%
RSE	9.89%	7.9%

TABLE XII
% IMPROVEMENT OVER RV CELL FOR RESIZED CELLS UNDER I_{trip}/I_{read} TEST

Cell	Worst-Case I_{trip}/I_{read} % improvement	Mean I_{trip}/I_{read} % improvement
RLE	2.51%	5.04%
RSE	11.6%	15.37%

cells they were derived from, as well as to the RV cell.

1) Leakage: As expected, the leakage performance of the resized cells is better than that of the SLE and SSE cells. For the RLE cell the leakage reduction when holding a '1' remains unchanged at a 6.96X reduction relative to RV, but the leakage reduction when holding a '0' only slightly reduces from 69.5X to 57.9X. (The SLE cell's leakage reduction when holding a '0' was only 2.5X). When the RSE cell is holding a '0' the leakage reduction stays at 2.04X relative to RV, and when it is holding a '1' the leakage reduction only changes from 6.96X to 6.79X. This change is also minimal when compared to the SSE's leakage reduction of 1.91X.

2) Performance: Due to the increased size of the pull-down NMOS transistors, the resized cells have the potential of improving the read-access time of the cell. For the RLE cell the discharge time along BLB remains at a 61.1% increase over the RV cell's BLB discharge time, but the BL discharge time is now only 3.7% longer than the RV cell's discharge time. As noted previously, only the BL discharge time is important due to the timed read based on the new sense amplifier. For the RSE cell, the discharge time along the fast side of the cell, BL, does not change, but the discharge time along BLB is reduced from the SE cell's 61.7% increase over RV to a 49.2% increase over RV. This extra performance along BLB plays no important role in the cell's performance. As for the write times, RLE's write time increases to a 39% increase over RV's write time from LE's 35.95% increase. The RSE write time jumps to a 45% increase over RV's write times.

3) Stability: The stability analysis has also been performed on the resized cells for both the SNM test and I_{trip}/I_{read} test. Both resized cells perform better than the RV cell in the worst case, and under Monte-Carlo analysis for the SNM. These results can be seen in Table XI. Under the I_{trip}/I_{read} test, the RLE cell now performs better than RV both in the worst-case and on average. The increase in N1's size accomplishes the higher I_{trip}/I_{read}. The RSE cell's I_{trip}/I_{read} value also increases slightly under all tests, even surpassing the SE cell's I_{trip}/I_{read} value in the worst-case. With a larger pull-down transistor, the process variations do not have as much an effect on the RSE cell's stability. These results are shown in Table XII.

F. Stability at Different Supply Voltages

Another figure of merit for the different cells is their stability under different supply voltages. For the technology being used, the nominal supply voltage is 1.2V. Monte-Carlo analysis has been performed for the RV, LE, SLE, RLE, SE, SSE and RSE cells for supply voltages ranging from 0.75V to 1.6V, for which the mean SNM is shown in Figs. 17 and 18.

From the plot it can be seen that for voltages above 1.2V, LE, SLE and RLE improve their SNM advantage over the RV cell. With a higher V_{GS}, the difference in conductance between the pass-gate (N3) and pull-down (N2) transistors, which was the root cause of the low stability at 1.2V, diminishes. At higher voltages, the SNM of the SE and SSE cells starts to diminish just as the SNM of the RV but at a lower rate. The SNM of the RSE cell levels off at higher voltages.

With lower supply voltages, the SNM of the asymmetric cells starts to suffer. For the LE, SLE and RLE cells, the SNM decreases rapidly, but SLE's SNM remains comparable to that of RV, while RLE's SNM becomes comparable to that of LE's. This decrease in stability is caused by the difference in conductance between RV and HV transistors at low V_{GS}'s. Furthermore, at low V_{GS}, the extra conductance of the larger transistor in the RLE cell does not have a large effect since the transistor is not fully on. The SNM of SE, SSE and RSE also decreases, but not as fast as that of LE. Again, this decrease in SNM is due to the difference in conductance at low V_{GS}'s.

Based on [12] the voltage regulator and power distribution network in microprocessors must maintain the supply voltage to within ±5% of nominal. Therefore, the reduced stability at low voltages for the asymmetric cells might not be a big concern except perhaps during any chip testing that may need to be performed at low voltage.

The same tests were performed for the I_{trip}/I_{read} method with the result that the curves for all cells are much better-behaved. The SE and SSE cells have a near 24% advantage over the RV cell at 0.75V and a 8% advantage at 1.65V. The LE and SLE cells have approximately a 16% decrease in I_{trip}/I_{read} at 0.75V and are comparable at 1.65V to the RV cell. The resized cells behave slightly differently, with the RSE cell having an 11.7% improvement at 1.65V and a 32.2% improvement at 0.75V. The RLE cell has a 9.6% improvement at 1.65V and a 4% decrease at 0.75V.

III. SENSE-AMPLIFIER

A conventional sense amplifier, shown in Fig. 19(a), is not suitable in our design due to the slow access time when the cell is storing a '0.' To obtain fast read times regardless of the data value, a new sense amplifier has been designed and is shown in Fig. 19(b). The design of this sense amplifier is based loosely on MOS Current-Mode Logic (MCML) ideas presented in [13]. Compared to the conventional sense amp, the new sense amplifier has 4 additional transistors and an area increase of roughly 0.229 μm^2 or 14.4%.

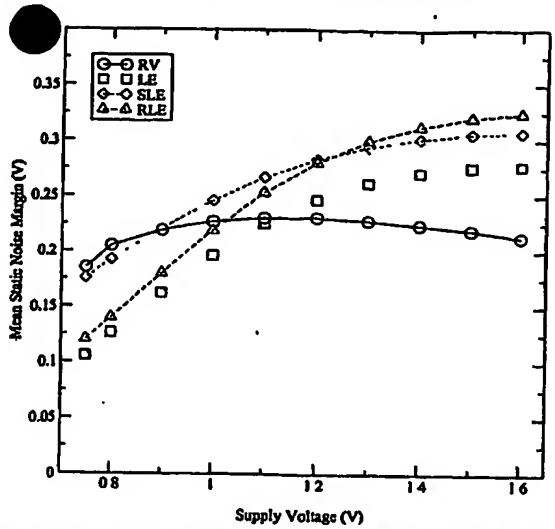

Fig. 17. Mean SNM under different supply voltage for LE derived cells

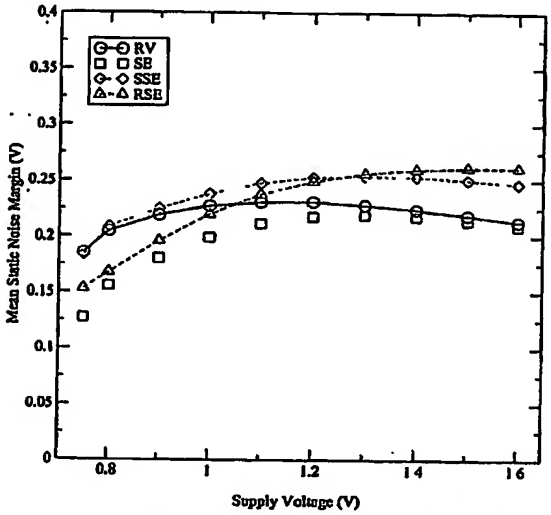

Fig. 18. Mean SNM under different supply voltage for SE derived cells

In addition to BL and BLB, the sense amp has two new inputs, D and DB. These are connected to a dummy column of cells that store '1' at all time, but which are otherwise exactly identical to all other cells in the array. This dummy column extends the full length of the SRAM array, so that, during every read operation, one of the dummy cells will have its wordline asserted. Since the dummy cells always store a '1', they are always fast on the discharge (as fast as the fast side of any other cell), and they are used to provide something like a timer signal. This is achieved by connecting the dummy bitlines to the sense amp in a reverse way (D connected to the right side, where BLB is connected, and DB connected to the left side, where BL is connected), so that D and DB trigger a fast read of a '0' result when the cell being read has '0' content.

Sensing a '1' is as fast as a conventional sense amp since this is done by sensing a discharge of BLB due to the action of the fast side of the cell. Sensing a '0' is *initiated* at a later time than it would be in a conventional sense amp. This is done to allow sufficient time for the fast side to trigger the sense amp if it has to do so. While initiating the sensing for a '0' is delayed,

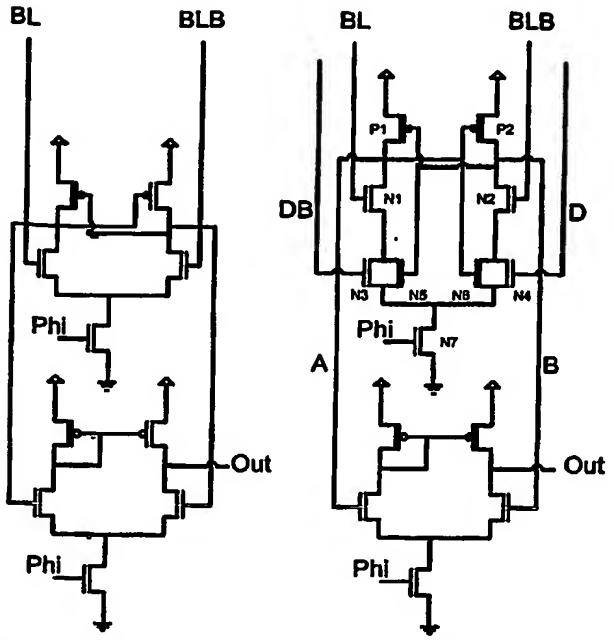

Fig. 19. (a) Conventional Sense Amplifier. (b) New Sense Amplifier

the combined effect of the dummy cell and the slow side of the asymmetric cell makes the sensing process itself much faster once initiated, so that the end result becomes available at about the same time as it would when sensing a '1'.

The detailed operation of the sense amplifier is as follows. Initially, the bitlines are precharged and all four amplifier inputs rise to V_{DD}. During this phase the sense amplifier is being reset and nodes A and B are reset to an intermediate value. During a read operation, either BLB will discharge (cell has a '1', fast discharge from the fast side) or BL will discharge (cell has a '0', slow discharge from the slow side). Furthermore the signal DB, which is on the fast side of the dummy cell, will be discharged since the dummy cells permanently hold a logic '1.' If BLB is being discharged (a logic '1' is being sensed), then the differential pair composed of N1 and N2 causes increased current to pass through the left branch, thus increasing the voltage at node B and decreasing the voltage at node A. Through the positive feedback loop of P1, P2, N5, and N6, the rate of change for nodes A and B are increased to achieve quick sensing. When BL is being discharged (a logic '0' is being sensed), then it does so at a slower rate since it is being discharged from the slow side of the asymmetric cell. To achieve fast sensing in this case also, the dummy bitlines, which are connected to the differential pair of N3 and N4, initiate the sensing of a logic '0.' Through the combined effect of the DB bitline being discharged and BL being discharged, albeit at a slower rate, approximately symmetric sense times are achieved.

For this sensing scheme to achieve reliable results it must allow for adequate time for BLB to discharge before initiating a logic '0' read. This safety factor is achieved in two ways. First the dummy bitlines are connected to all sense amps and therefore have a slightly higher capacitive load compared to real bitlines leading to a slower discharge on DB compared to BLB. The extra capacitive loading does not slow the sense time when

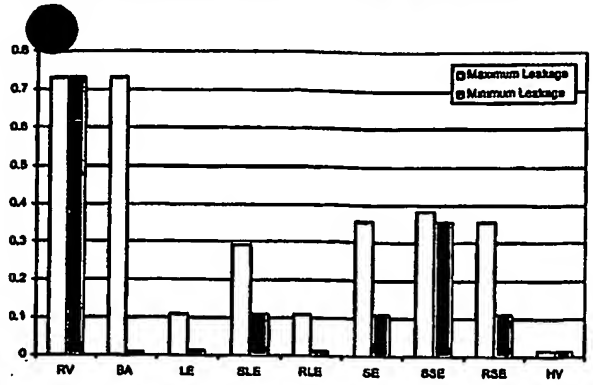

Fig. 20. Maximum and Minimum Leakage Current Attributable to Cells

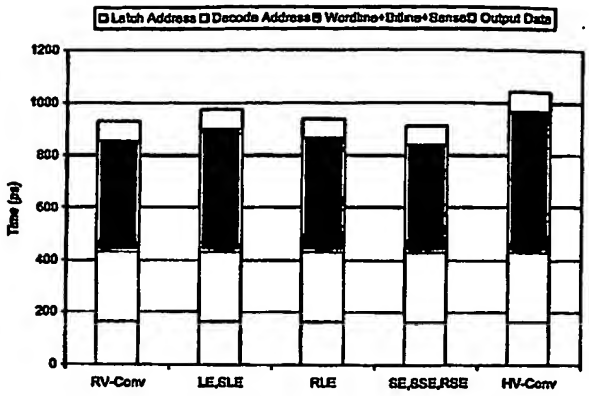

Fig. 21. Breakdown of Memory Access Time

BL is discharging because of the concerted effort between BL and DB to sense the same value. Second the transistors connected to the bitlines are wider than the transistors connected to the dummy bitlines leading to a higher transconductance. This leads to a higher gain from the bitlines to the output than from the dummy bitlines. We have also performed sensitivity analysis of this sense amplifier, and it performs on par with the conventional sense amplifier.

To limit the sense power, the sense amplifiers are clocked as in [14][15][16]. The sense clock turns on the amplifiers and sets them up in their high gain region before the sensing occurs. To improve yield and ensure low-power operation, the clock path must be matched to the data path. This matching is achieved by using an extra set of dummy bitlines to match the bitline delay and clock the sense amplifiers at the appropriate time as in [16].

IV. SRAM

Using the above cells and the sense amplifier presented above, a 32-Kbyte SRAM was designed and simulated to measure leakage, and read and write times. Each of the 128 SRAM sub-arrays contains 64 cells along each bitline, and 32 cells along each wordline. The SRAM was simulated at a temperature of 110°C with the RV, BA, LE, SLE, RLE, SE, SSE, RSE and HV cells. Furthermore, the RV and HV cells were simulated with a conventional sense amp, and these results were used as a reference for our design.

Fig. 20 shows the total leakage within the SRAM attributable to the SRAM cells when the SRAM is either holding all '0's or all '1's. The leakage includes the leakage needed for the dummy cells (their contribution is negligible, however, given the size of the SRAM). The leakage trends seen above for the single cell remain true for the complete SRAM, where LE and SE offer a reduction of 70X and 2X while storing a '0' and a reduction of about 7X when storing a '1.' The stability improved cells, and the resized cells also show the same leakage trends from the single cell experiments.

The total SRAM read access time includes four components: 1) input register propagation delay and hold times, 2) the address decoding delay, 3) the delay for wordline, bitline and sensing, and 4) the output register setup time. Our simulation results showing these components for the various SRAM arrays are shown in Fig. 21. Notice that only the 3rd component is affected by the cell design. Specifically, this time is the time period from when precharging is complete to when the sense amplifier has reached 90% of its swing.

Fig. 22 shows the sense times (the 3rd component of Fig. 21) for all the cells (SLE, SSE and RSE are not shown for clarity, because their sense times are similar to the sense times of LE, SE and SE, respectively). While the discharge times are asymmetric, it can be seen that the worst-case sensing times are now on-par with the RV cell with a conventional sense amplifier. Compared with the RV cell with a conventional sense amp, the LE cell is 10% slower (although the effect on the total read time is an increase of just under 5%, as seen in Fig. 21), but the SE cell is slightly faster (note this is not because the sense amplifier is quicker, but because the bitline discharge time for the SE cell is 50ps quicker than that of the RV cell, which is a by-product of the asymmetry of the SE cell). Furthermore, the RLE cell has a worst-case sense time that is 2.5% slower than the RV cell, with the effect on total read time being near 1%. Interestingly, the HV cell with a conventional sense amplifier would be 26% slower.

An important side comment to be made is that the new sense amplifier does not speed up the sensing for RV and HV when compared to the sensing with the conventional sense amplifier. Indeed, the RV and HV cells with the new sense amplifier have worst-case sense times which are 5% slower than the sense times with the conventional sense amplifier. Thus, in comparing the speed of the new cells with the new sense amp to the conventional cells with the conventional sense amp, the comparison is *fair* and valid, because the new sense amplifier on it's own does not speed up the read access time of the conventional cells.

Finally, the write times for the different cells are shown in Fig. 23. The LE and SE cells show an increase of 19.4% and 25.3% respectively over the RV cell. The SLE and SSE cells show an increase of 28.4% and 13.4% respectively, and finally the RLE and SLE show an increase of 22.4% and 27.6% respectively. The increase in write times is of minor importance since the write times are all shorter than the read times of the associated cells and therefore the speed of the SRAM is dependent on the read time.

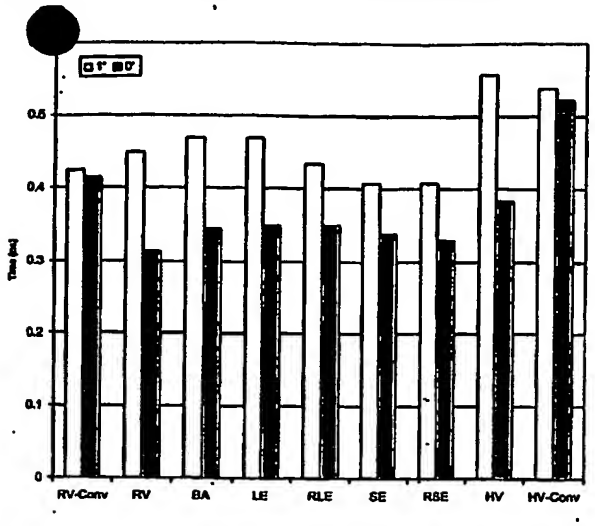

Fig. 22: Sense times during a read cycle, i.e., the 3rd component of Fig. 21.

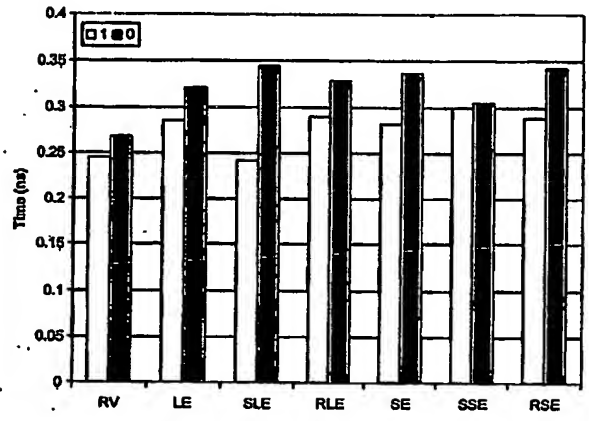

Fig. 23. Write Times

V. ARCHITECTURAL ENHANCEMENTS

We investigated two cache organizations that use asymmetric cell designs: *statically biased* and *dynamic inversion*. In the *statically biased* cache, the cells are simply replaced with asymmetric ones. This cache is *statically biased* to dissipate low leakage power only when it stores the preferred bit value ('0'). What makes this cache successful is *typical* program behavior: as we show in [8], the SPEC2000 programs we studied exhibit a strong bias towards zero. Specifically, we observed that a level-1 data cache had an average 78.7% zeros in the data stream, and a level-1 instruction cache had an average of 62.9% zeros. Given this, the statically biased cache with the SE cells reduces leakage by 4.5X and 3.8X for an instruction and a data cache, respectively, compared to conventional symmetric-cell caches. The caches are 32Kbyte 4-way set associative caches. While programs with a higher fraction of '1's than '0's may exist, our SRAM would still dissipate much lower leakage power compared to the regular-V_t cell cache.

In *selective inversion*, the values stored within a block can be inverted at a byte granularity (other granularities are possible). In this design, if a byte contains five or more ones it is inverted prior to storing it in the cache. This cache needs an additional *inversion flag* cell per byte that holds information on which bytes were inverted. Inversion happens at write time. Since stores are typically buffered in a write buffer and are only sent to the data cache on commit, there is plenty of time to decide and apply inversion if necessary. Additional area, dynamic power and performance trade-offs apply to this design. An investigation of these issues is part of our ongoing and future work.

VI. CONCLUSION

In this paper, we proposed a novel approach that combines both circuit- and architecture-level techniques. Our approach drastically reduces leakage power dissipation. The key observations behind our approach are that cache-resident memory values of ordinary programs exhibit a strong bias towards zero or one at the bit level.

We introduced a family of high-speed asymmetric dual-V_t SRAM cell designs that exploit this bit-level bias to reduce leakage power while maintaining high performance. The six best asymmetric cells offer different performance/leakage/stability characteristics. The SE cell reduces leakage power by at least 2X and by 7X in the preferred state. It is as fast as the conventional, RV, SRAM cell. By comparison, the LE cell reduces leakage by at least 7X and by about 70X in the preferred state. Its total read time is only 5% higher than the SE and RV cells. These latter two cells have lower stability than LE under both the SNM *and* the I_{trip}/I_{read} tests. Four other cells that compensate for stability were designed, two by choosing different combinations of threshold voltages for the cell transistors, and two by changing some transistor sizes. The SSE cell reduces leakage power by 1.9X and 2.3X in the preferred state with no performance degradation, and the SLE cell reduces leakage power by 2.3X and 7X in the preferred state with only a 5% increase in read access times. The SSE and SLE cells have comparable stability to the RV cell. The RLE cell reduces leakage by 58X in the preferred state and by 7X in the other state with only a 1% increase in read access time, and an area increase of about 2.4%. The RSE cell reduces leakage by about 7X in the preferred state, and 2X in the other state. It has no performance degradation, but has an area increase of about 2.9%. The RLE and RSE cells have comparable stability to the RV cell. By comparison, an all high-V_t cell reduces leakage power by about 70X while its bitline discharge time is 60% slower than the SE and RV cells.

We also proposed two cache organizations that used either a static bias towards zero, or dynamic, selective inversion to maximize the number of cache bits that are zero. While the reduction possible with either technique depends on application behavior, for the SPEC2000 benchmarks which we considered, the statically biased cache with the SE cells reduces leakage by 4.5X and 3.8X for an instruction and a data cache, respectively, compared to conventional symmetric-cell caches.

A summary of the results pertaining to all the cells, relative to the RV cell, is shown in Table XIII. Here, "Leakage (0)" and "Leakage (1)" refer to the leakage when the cell is storing a '0' and a '1', respectively, "Delay" refers to the total read access time, and "Area" refers to the total cell layout area. If we work with the observed average of about 70% 0s and 30% 1s, we can

TABLE XIII

SUMMARY RESULTS FOR ALL THE CELLS.

Cell	Leakage (0)	Leakage (1)	Δ Delay	Δ Stability (SNM)	Δ Stability (I_{trip}/I_{read})	Δ Area
RV	100%	100%	0%	0%	0%	0%
LE	1%	14%	5%	7%	-5%	0%
SE	14%	50%	0%	-6%	15%	0%
SLE	14%	43%	5%	23%	-7%	0%
SSE	43%	53%	0%	9%	13%	0%
RLE	2%	14%	1%	22%	5%	2%
RSE	15%	49%	0%	8%	15%	3%

TABLE XIV

SUMMARY RESULTS FOR ALL THE CELLS, SHOWING EXPECTED LEAKAGE.

Cell	Expected Leakage	Δ Delay	Worst Δ Stability	Δ Area
RV	100%	0%	0%	0%
LE	5%	5%	-5%	0%
SE	25%	0%	-6%	0%
SLE	23%	5%	-7%	0%
SSE	46%	0%	9%	0%
RLE	6%	1%	5%	2%
RSE	25%	0%	8%	3%

give projections for the "Expected Leakage" for the long-term average leakage of an SRAM array (accounting only for the cell leakage), as shown in Table XIV, where the column for "Worst Δ Stability" gives the worst case between the two columns of Table XIII corresponding to the SNM test and the I_{trip}/I_{read} test. If one had to single out *the best* cases, it is perhaps the case that SSE and RLE combine the best features. SSE has less than half the original leakage of the RV cell with no loss of either performance, stability (mean, at nominal voltage) or area. RLE has only 6% of the original leakage of the RV cell with a performance loss of only 1%, no loss of stability, and an area increase of only 2%.

REFERENCES

[1] Ş. Borkar. Design challenges of technology scaling. *IEEE MICRO*, 19(4):23–29, July–August 1999.

[2] T. Kam et. al. Eda challenges facing future microprocessor design. *IEEE Transactions on Computer-Aided Design*, 19(12), December 2000.

[3] F. Hamzaoglu et. al. Dual v_t-sram cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13um technology generation. *Proceedings of the 2000 International Symposium on Low Power Electronics and Design*, July 2000.

[4] Ş. Kaxiras, Z. Hu, and M. Martonosi. Cache decay exploiting generational behavior to reduce leakage power. *Proceedings of the 27th International Symposium on Computer Architecture*, July 2001.

[5] S.-H. Yang, M. D. Powell, B. Falsafi, K. Roy, and T. N. Vijaykumar. An integrated circuit/architecture approach to reducing leakage in deep-submicron high-performance i-caches. *Proceedings of the 7th International Symposium on High-Performance Computer Architecture*, January 2001.

[6] H. Zhou, M. C. Toburen, E. Rotenberg, and T. M. Conte. Adaptive mode control: A static-power-efficient cache design. *Proceedings of the 2001 International Conference on Parallel Architectures and Compilation Techniques*, September 2001.

[7] John L. Hennessy and David A. Patterson. *Computer Architecture: A Quantitative Approach*. Morgan Kaufman, 2 edition, 1996.

[8] Navid Azizi, Andreas Moshovos, and Farid N. Najm. Asymmetric-cell caches: exploiting bit value biases to reduce leakage powerin deep-submicron, high-performance caches. ECE Computer Group TR-01-01-02, University of Toronto, 2002.

[9] E. Seevinck Sr., F. J. List, and J. Lohstroh. Static-noise margin analysis of mos sram cells. *IEEE Journal of Solid-State Circuits*, 22:748–754, October 1987.

[10] A. Bhavnagarwala, X. Tang, and J. Meindl. The impact of intrinsic device fluctuations on cmos sram cell stability. *IEEE Journal of Solid-State Circuits*, 36, April 2001.

[11] E. Ajith Amerasekera and Farid N. Najm. *Failure Mechanisms in Semiconductor Devices*. John Wiles and Sons, 2 edition, 1997.

[12] Ed Grochowski, Dave Ayers, and Vivek Tiwari. Microarchitectural simulation and control of di/dt-induced power supply voltage variation. *Proceedings of the Eighth International Symposium on High-Performance Computer Architecture*, 2002.

[13] J. M. Musicer and J. Rabaey. Mos current mode logic for low power, low noise cordic computation in mixed-signal environments. *Proceedings of the 2000 International Symposium on Low Power Electronics and Design*, July 2000.

[14] T. Hirose et al. A 20ns 4mb cmos sram with hierarchical word decoding architecture. *IEEE International Solid-State Circuits Conferance., Digital Technical Papers*, pages 132–133, 1990.

[15] K. Itoh, K. Sasaki, and Y. Nakagome. Trends in low-power ram circuit technologies. *Proceedings of the IEEE*, 83(4), April 1995.

[16] B. S. Amrutur and M. A. Horowitz. A replica technique for wordline and sense control in low-power sram's. *IEEE Journal of Solid-State Circuits*, 33:1208–1219, August 1998.